

Unsupervised Image Segmentation Using a Hierarchical Clustering Selection Process*

Adolfo Martínez-Usó, Filiberto Pla, and Pedro García-Sevilla

Dept. Lenguajes y Sistemas Informáticos, Jaume I University
{auso, pla, pgarcia}@uji.es
<http://www.vision.uji.es>

Abstract. In this paper we present an unsupervised algorithm to select the most adequate grouping of regions in an image using a hierarchical clustering scheme. Then, we introduce an optimisation approach for the whole process. The grouping method presented is based on the maximisation of a measure that represents the perceptual decision. The whole strategy takes profit from a hierarchical clustering to find a maximum of the proposed criterion. The algorithm has been used to segment real images as well as multispectral images achieving very accurate results on this task.

1 Introduction

Image segmentation has been a very focused topic in the literature. Looking for the main regions that perceptually compose an image has been the target for many researchers [10,3,5,8]. This task has been used as an aim in itself or as a preprocessing step. In both cases, the difficulties involved in this process are very well-known [4].

The main motivation of this work has been to obtain a robust and completely unsupervised segmentation criterion based on perceptual similarities among the image regions and the contour information. Thus, on one hand, the method will be guided by a merging process using an agglomerative hierarchical clustering and, on the other hand, the method will satisfactorily select the optimal, or sub-optimal, partition in this hierarchical structure combining region and boundary information.

Thus, the method presented is based on two basic steps which define our segmentation strategy:

- A **hierarchical clustering**, in order to create a structure of non-overlapping partitions from an oversegmented representation until the stage with just one region (cluster). The hierarchical structure will serve as the guide to an optimisation strategy. It is supposed that the hierarchical structure will contain the optimal (or close to optimal) segmentation.
- A **global similarity measure** as a criterion function to be optimised, which is applied through the above named structure looking for a partition that optimises the criterion function. This partition will represent the final segmentation result.

* This work has been partly supported by projects ESP2005-07724-C05-05 from Spanish CI-CYT and P1-1B2004-36 from Fundació Caixa-Castelló.

We apply our algorithm not only to classical real images from the literature, which is useful to test the performance of the method, but especially to multispectral images with more than three bands as well. For instance, in some applications, like fruit quality inspection tasks, there exist some types of defects that can only be detected in certain bands of the spectrum, and most of the defects have a specific range of bands where they can be better discriminated. Therefore, for some inspection tools, multispectral images are becoming useful to achieve better recognition and classification results. These results regard to an application in which we are working on as a part of a quality fruit estimation project on oranges.

The whole segmentation process is summarised in the following algorithm. It is particularly interesting to emphasise the two final steps of the algorithm where the most relevant contribution of this work is.

1. Preliminary steps on cluster initialisation (Sect. 2).
2. The method compares clusters and iteratively merges neighbouring regions to create a structure that constitutes a hierarchical family of derived clusters (Sect. 3).
3. One of those previous partitions will maximise a criterion function used to choose the right clusters, being the chosen partition the final segmentation result (Sect. 4).

Finally, many segmentation techniques have high computational requirements, especially the iterative ones. These requirements become higher when the image size increases. Thus, a study of the evolution of the distances among the image regions is also presented (see Sect. 5). This study will allow us to introduce an important optimisation in the process without making worse the final segmentation results.

2 Preliminary Steps for Clustering Initialisation

This phase of the algorithm identifies the main homogeneous regions in the image. This initialisation uses a quadtree structure (QT) to represent the image, which is a multiresolution representation commonly used to split or decompose a given image into similar square regions. The QT also serves to represent the spatial relationships among regions.

First merging, while the QT is growing up: The hierarchical structure developed by a QT has several levels of resolution made by nodes and leaves. Every time a level is developed, the algorithm checks if any two of the current leaves can be merged. Because of the spatial information contained in the QT representation, neighbourhood operations are completely defined from the QT structure [12].

Second merging, look for high gradient magnitudes: Every region formed in the previous step starts growing up while the gradient magnitude increases at the region border. Thus, each region border grows towards the high gradient magnitude values. The gradient magnitude considered for each pixel is the maximum gradient value found in all bands.

This behaviour is usually formulated as an active contour functional [9], whose internal energy has the discrete form $E_i = -\sum_{x \in \beta_i} |\nabla g(x)|$, where, β_i represents the set of boundary pixels in region i and the value $|\nabla g(x)|$ returns the gradient magnitude

at pixel x . Regions previously formed are used as the active contours initialisation. Of course, on this functional, the energy E in each contour region has to be minimised. It is important to note that if two regions compete for a third region, which is neighbour of both of them, the algorithm computes the average of the gradient magnitudes of each region border, and merges the regions with the highest difference.

Smoothing regions: After the previous steps, regions very often present isolated small holes inside of them, that is, many times there is a gradient magnitude peak due to impulsive or other type of noise. Thus, this step looks for these kind of isolated regions and merges them with the larger regions which contain the isolated ones.

3 Hierarchical Clustering in the Image Domain

Some relevant recent works for image segmentation are based on clustering or grouping processes [6,7]. These methods are designed to discover and extract hidden structures in data sets [2]. Regarding to the procedure we use, it starts with as many clusters as the initialisation step provides. From these initial clusters, the method forces a merging operation in each iteration, that is, the algorithm always eliminates a cluster in each step. This means that each iteration compares every pair of spatially connected clusters and calculates the distance (D) between them. The two clusters, that is, image regions, with more similar values are forced to merge. Distance D is defined as follows:

$$D_{ij} = d_{ij}(1 + \delta_{ij}) \quad (1)$$

where,

$$d_{ij} = (\mu_i - \mu_j)^T (\Sigma_i^{-1} + \Sigma_j^{-1}) (\mu_i - \mu_j) \quad \delta_{ij} = \frac{\sum_{x \in \beta_{ij}} |\nabla g(x)|^2}{\|\beta_{ij}\|} \quad (2)$$

Mean values on the intensity of clusters i and j are represented by μ_i, μ_j whereas their covariance matrices are represented by Σ_i, Σ_j . In δ_{ij} function, β_{ij} is the set of pixels that belong to the boundary between regions i and j . Function $\|\cdot\|$ returns the cardinality of a set.

Note that the proposed measure D_{ij} can be evaluated analytically as a mixture of two expressions. Under the assumption of considering normally distributed clusters, d_{ij} value is a Mahalanobis distance between two distributions, defined in the pixel domain (gray level, colour, multispectral, ...). This term in the distance expression D accounts for the similarity in pixel values of the distribution of pixel values in both regions considered to be merged. On the other hand, δ_{ij} averages the gradient magnitude values on the boundary between these two connected clusters. Thus, this term accounts for the strength of the discontinuity between the distributions of pixel values between the considered clusters, including a spatial measure of discontinuity in the distance function.

4 Clustering Assessment

While the algorithm completes the process described in the previous section, a non-parametric estimation of the goodness of the data partition is performed. That is, the

final result is selected without any a priori information about the final number of clusters or the shape of the resulting regions. An example of this type of approach can be found in [11] which is very similar to ours. However, their method suffers from problems on images where the boundary information has special relevance.

The criteria to select an optimal clustering from the previous hierarchical structure is completely independent of the way in which the hierarchical tree structure is constructed using the distance measure introduced in the previous section.

The algorithm has been formulated as the maximisation of a criterion function S that represents, given an image partition, how well the pixels fit to the corresponding regions. In such a way, it maximises the following expression:

$$S = S_i \cdot S_e \quad (3)$$

being S_i a inner measure and S_e a external measure of each partition in the hierarchical tree structure. S_i can be considered as an average measure that a pixel in the image belongs to the region it has been assigned to, and S_e an average measure that a pixel does not belong to its neighbouring regions. Therefore, the criterion function S takes into account that the pixels in the image belong to the assigned regions in the partition and, simultaneously, that the pixels do not belong to neighbouring regions in the image domain. This can be considered as a perceptual measure of the grouping, in such a way that it estimates how "well" the pixels are grouped and they are consistent internally and, at the same time, different enough from spatially nearby clusters.

Given a particular partition, the inner and external average pixel measures, S_i and S_e respectively, are defined as follows:

$$S_i = \frac{1}{N} \sum_R \sum_{x \in R} S(x, R) = \frac{1}{N} \sum_R \sum_{x \in R} e^{\frac{-(x - \mu_R)^2}{2\sigma^2}} \quad (4)$$

$$\begin{aligned} S_e &= \frac{1}{N} \sum_R \sum_{x \in R} S(x, NR(R)) = \frac{1}{N} \sum_R \sum_{x \in R} \prod_{R' \in NR(R)} S(x, R') = \\ &= \frac{1}{N} \sum_R \sum_{x \in R} \prod_{R' \in NR(R)} (1 - S(x, R')) = \frac{1}{N} \sum_R \sum_{x \in R} \prod_{R' \in NR(R)} \left(1 - e^{\frac{-(x - \mu_{R'})^2}{2\sigma^2}} \right) \end{aligned} \quad (5)$$

In these equations, N represents the total number of pixels in the image. R , R' are regions and $NR(R)$ is the set of neighbouring regions of region R . The pixel value is represented by x whereas μ_R is the average value of the pixels in the region R . $S(x, R)$ is a similarity measure between pixel x and region R .

Equation (4) represents the sum of the inner values for each cluster. That is, it expresses a measure that image pixels belong to its current assigned cluster. This is estimated assuming a Gaussian distribution of region pixel values, characterised by a mean and an expected variance. The expected variance is fixed and it is a smoothing parameter of the expected segmentation.

On the other hand, equation (5) provides the external values. It expresses a measure that image pixels do not belong to its current neighbouring clusters. This is also estimated assuming a Gaussian distribution of the pixel values of the neighbouring clusters.

The way this value is expressed is the complementary of the measure that a pixel belongs to the neighbouring regions $S(x, NR(R))$, which is defined as the measure that a pixel does not belong to any of the neighbouring regions, that is, $S(x, NR(R)) = \prod_{R'}^{NR(R)} (1 - S(x, R'))$.

In these equations σ is always a fixed value representing the variance threshold we permit on the density-estimation. This variance represents a smoothing parameter of the expected segmentation. The lower σ , the more clusters will have the chosen partition using function (3). Nevertheless, in the method here described, once the σ was selected experimentally, it was fixed to the same for all the experiments.

5 Optimising the Merging Process

It is a well-known fact that the main drawback of iterative methods is how much time/resources the algorithm needs to perform these iterations. The problem becomes worse when the algorithm has to deal with complex structures and the process advances bit by bit due to the need to check the clustering robustness. In this sense, we can try to optimise the process of merging regions.

Graphs in Fig.1 show the distance values between any pair of neighbouring regions arranged by this value for two different images at three different stages of the process. That is, distances calculated as described in section 3 are arranged from lower to higher value. Each time two regions are merged, the algorithm has to recalculate the affected distances. Fig. 1 shows the number of regions NR to merge under each graph. The x-axis corresponds to the initial number of distances computed for the corresponding image and the y-axis represents the distance value.

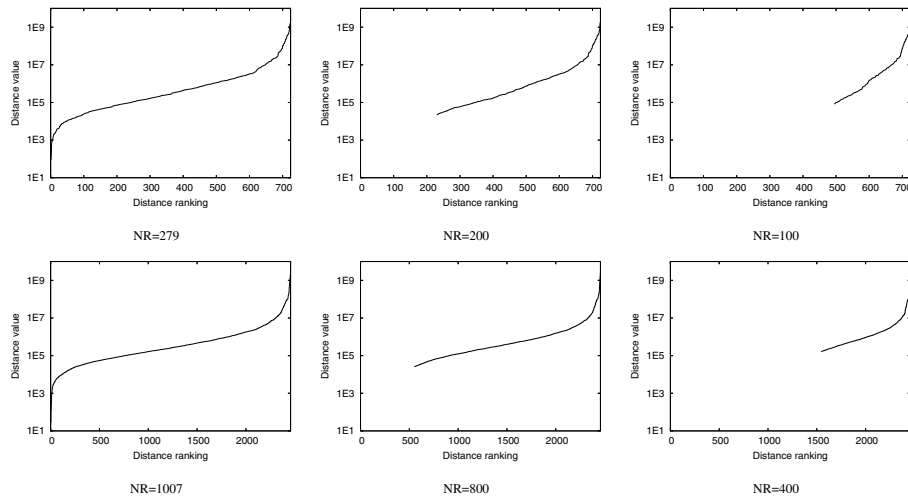


Fig. 1. Distance evolution examples. First row, toys RGB image. Second row, orange multispectral image. Coordinate axes: number of distances (x) and distance values (y). NR gives the initial number of regions considered.

As we can see from the graphical results shown on this figure, the distance evolution presents a very regular behaviour and, therefore, a very predictable evolution. In such cases, it is clearly desirable to reduce as much as possible this phase in order to take profit of the particular nature of this evolution. In this sense, instead of merging only two clusters at each iteration, we start from the pair of clusters with the minimum value of D and move to the pair of clusters which provided the next smallest distance value if both clusters has not been previously merged at this iteration. We keep merging pairs of clusters according to their distance value until the distance value reaches a threshold which is established depending on the maximum distance value.

6 Results

Maximising the inner and external average measures product for all partitions considered gives us the final segmentation result. Fig. 2 shows two graphical results¹ drawing the behaviour of criterion function (3). The graph axes represent the number of clusters (x-axis) and the corresponding S value of the partition (y-axis). First graph in Fig. 2 has been obtained for the *toys* image, where the maximum value is reached at 25 clusters. Next graph, obtained from the multispectral image of an *orange* (the one on the first row in Fig. 4), has its maximum value at 8 clusters. The segmentation results for these images can be seen in figures 3 and 4. In figure 3, other three results are shown for classical images as *peppers*, *tree* and *beans*. As we can see, the results in all the images are consistent with a perceptual interpretation, with well defined contours and all important regions detected. Note that the results are presented using an edge image where the darker the edge, the greater difference between mean values of neighbouring regions.

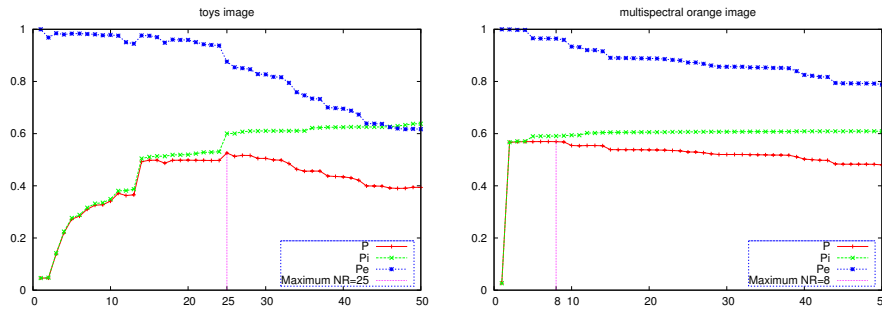


Fig. 2. Maximum likelihood estimation. On the left a real image (*toys*), on the right a multispectral image of an orange.

Although this algorithm has achieved satisfactory results on gray level and RGB colour images, it is expected that with more complex images, like multispectral images, it may also provide consistent results. In this sense, a collection of multispectral images

¹ In all the graphical results, we have marked on the x-axis the position which provided the maximum value of S .

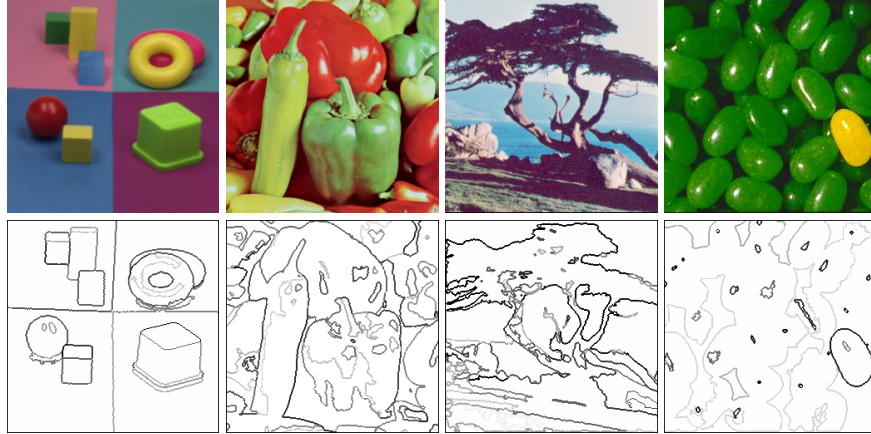


Fig. 3. Examples of results on classical images. From left to right, toys, peppers, tree and beans.

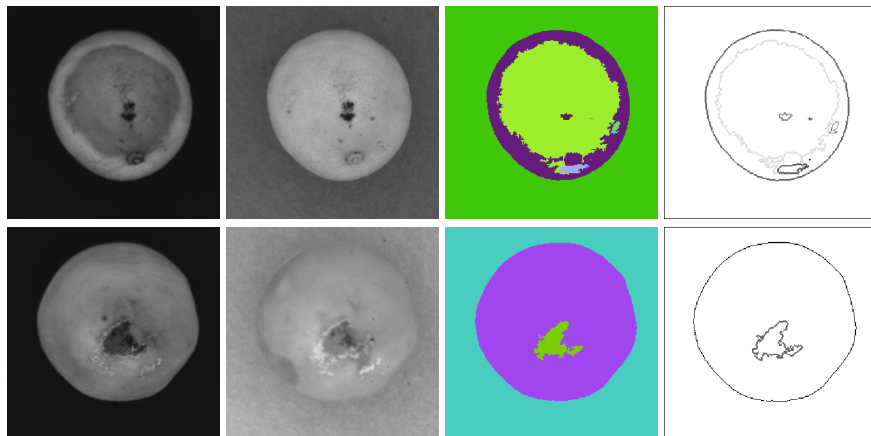


Fig. 4. Multispectral images of oranges. Two first columns show two example bands from each orange (420/640 nm and 680/1000 nm). The following columns show the segmentation results with random colours and the edges for the same results.

of oranges obtained by an imaging spectrograph (RetigaEx, Opto-knowledge Systems Inc., Canada) has been used. The spectral range is extended from 400 to 1050 nm with a spectral resolution of 10 nm for each band. The database used includes several kinds of orange defects. Fig.4 shows the result of applying the algorithm on two of these images. The orange on the top presents overripe and scratch defects, whereas the orange on the bottom suffers from a rotten defect. Note that it has correctly segmented the orange parts, labelling defect skin and healthy skin as different regions. We can also see the graph (Fig.2) obtained for the first orange image where, according with the results shown in Fig.4, the maximum value is at 8 regions. The average values S_i and S_e can also indicate the accuracy of the segmentation results. That is, low values in

Table 1. The "/" character separates values for original segmentation / values for optimised segmentation. NR column presents the final number of regions. S is the final value considered. Finally, last column gives the Borsotti value for each segmentation.

image	NR	S	Borsotti value
beans	38/38	0.2662/0.2661	11886.23/11884.31
toys	25/25	0.5258/0.5258	719.18/721.27
peppers	88/90	0.2931/0.2674	3155.00/4510.90
tree	100/95	0.2861/0.2633	7797.52/7425.97
orange (top)	8/6	0.5698/0.5802	1795.00/1578.76
orange (bottom)	3/3	0.4122/0.4125	2689.28/2689.28

these parameters may result on poor segmentation images, probably, because the input images are corrupted or affected by noise. Empirically, results where $S_i \cdot S_e \leq 0.2$ are discarded. As we can see on graphs from Fig. 2, the S value is noticeable bigger than this value, indicating the reliability of the resulting segmentations.

In addition, we also present the improved results derived from the use of the optimisation proposed in section 5. For our experimental evaluation purposes, and due to the fact that differences between the original segmentation results and the optimised ones are not clearly visible (they seem identical), quantitative results for the images used on this paper are given in table 1, demonstrating the performance of the proposed optimisation. In [1], Borsotti et al. proposed a measure to estimate the quality of the segmentation results. This value measures the intra-region uniformity and the inter-region contrast. It also has a penalisation factor inversely proportional to the number of regions in segmented images [13]. Thus, we also present this well-known value on table 1 as an indicator of how close the two segmentation results are².

Finally, it is important to stress that the final number of regions (based on the number of clusters) is determined by the algorithm with no prior knowledge about the image. The σ value used is the same in all the experiments carried out for this work. In this case, $\sigma = \sqrt{10}$.

7 Conclusions

An unsupervised image segmentation algorithm has been presented. It performs the clustering in the image domain, using spatial information to build a hierarchical clustering structure. The hierarchical clustering is performed using a proposed distance function that tries to integrate the similarity in the distribution of pixel values, and the edge information, in order to adapt region borders to image edges. In our experiments, this process has been improved by a significant optimisation in terms of speed and without affecting the solution quality. Moreover, a criterion function based on the maximisation of a similarity measure has been introduced. This function estimates the right number of clusters in the hierarchical structure, in such a way that it is in accordance with a perceptual decision about the grouping.

² Note that we had to develop an extension of Borsotti quality estimation equation to multispectral values.

The proposed method has been tested in gray level, colour and multispectral images, providing satisfactory results in the chosen groupings and the segmentation results, where regions are well defined by contour edges.

Although the criterion function used behaves as expected, further work is directed to achieve a smoother criterion function that can avoid local oscillations and, therefore, trying not to get stuck in local maxima around the global maximum, in order to achieve a more accurate decision about the final grouping representing the segmentation.

References

1. M. Borsotti, P. Campadelli, and R. Schettini. Quantitative evaluation of color image segmentation results. *Pattern Recognition Letters*, 19:741–747, 1998.
2. Joachim Buhmann. Data clustering and learning. *The Handbook of Brain Theory and Neural Networks*, 2nd Edition:308–312, 2002.
3. Heng-Da. Cheng, X.H. Jiang, Ying Sun, and Jingli Wang. Color image segmentation: Advances and prospects. *Pattern Recognition*, 34(12):2259–2281, 2001.
4. E.M. Gurari and H. Wechsler. On the difficulties involved in the segmentation of pictures. *PAMI*(4), (3):304–306, 1982.
5. R.H. Haralick and L.G. Shapiro. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, (29):100–132, 1985.
6. J. Keuchel, M. Heiler, and C. Schnörr. Hierarchical image segmentation based on semidefinite programming. 3175:120–128, 2004.
7. S. Makrogiannis, G. Economou, S. Fotopoulos, and G.B. Nikolaos. Segmentation of color images using multiscale clustering and graph theoretic region synthesis. *IEEE Transactions on Systems, Man, and Cybernetics*, 35(2):224–238, 2005.
8. X. Muñoz, J. Freixenet, X. Cufí, and J. Martí. Strategies for image segmentation combining region and boundary information. *PRL*, 24(Issue 1-3):375–392, 2003.
9. Mark Nixon and Alberto S. Aguado. *Feature Extraction in Computer Vision and Image Processing*. 2002.
10. N.R. Pal and S.K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, 1993.
11. E.J. Pauwels and G. Frederix. Finding salient regions in images: Non-parametric clustering for image segmentation and grouping. *Computer Vision and Image Understanding*, 75(1/2):73–85, 1999.
12. H. Samet. *Applications of Spatial Data Structures: Computer Graphics, Image Processing and GIS*. 1990.
13. Yu Jin Zhang. A review of recent evaluation methods for image segmentation. *Proceedings of the Sixth ISSPA*, 1:148–151, 2001.