ON BUILDING A HIERARCHICAL REGION-BASED REPRESENTATION FOR GENERIC IMAGE ANALYSIS

Veronica Vilaplana, Ferran Marques

Technical University of Catalonia (UPC), Barcelona, Spain
{veronica, ferran}@gps.tsc.upc.edu

ABSTRACT

This paper studies the procedure to create a hierarchical region-based image representation aiming at generic image analysis. This study is carried out in the context of bottom-up segmentation algorithms and, specifically, using the Binary Partition Tree implementation. The different steps necessary to create a hierarchical region-based representation are analyzed; namely, (i) the creation of the initial partition in the hierarchy, which is split into the definition of the initial merging criterion and the proposal of a stopping criterion, and (ii) the merging criteria used to produce the different regions in the final hierarchical representation. For both steps, the proposed approach is assessed and compared with previous existing ones over a large data set using well-established partition-based metrics.

Index Terms— Image segmentation, Object detection, Image analysis

1. INTRODUCTION

Image segmentation is a basic initial step for a large variety of applications [1]. The large scope of these applications, jointly with the fact that image segmentation is an ill-posed problem, has resulted in a proliferation of specific segmentation techniques aiming at solving concrete problems. Nevertheless, the possibility of having a generic segmentation approach that would provide a good/high quality initial point for subsequent specific analysis is still necessary, both as starting step for dealing with particular scenarios and as basic tool for applications inherently dealing with generic images (e.g.: semantic indexing of generic image databases).

One approach to region-based generic image analysis is to not constrain the image representation to a single partition but to create a hierarchy of partitions representing the image at different resolution levels [1]. The idea is to have a universe of partitions representing the image at various resolutions, out of which a more specific algorithm can select the most convenient region(s) for its concrete application. The selected region(s) may represent objects or good object's estimations which could, in turn, launch a refining process.

Among the existing hierarchical representations, the Binary Partition Tree (BPT) [2] proposes a hierarchy in terms of regions, in contraposition to those techniques that propose a hierarchy of partitions. The BPT representation is based on a region merging algorithm. Starting from an initial partition (that can be as fine as assuming each pixel is a region), the region merging algorithm proceeds iteratively by (1) computing a similarity measure (*merging criterion*) for all pair of neighbor regions, (2) selecting the most similar pair of regions and merging them into a new region and (3) updating the neighborhood and the similarity measures.

The BPT stores the whole merging sequence from the initial partition to the one-single region representation. Leaves in the tree are the regions in the initial partition. A merging is represented by creating a parent node (the new region resulting from the merging) and linking it to its two children nodes (the pair of regions that are merged). An example of BPT is shown in Figure 6. In spite of its versatility, some problems arise in order to build a hierarchical representation. These are the problems that are analyzed in this paper:

Definition of the initial partition in the hierarchy: The initial partition is not only necessary to reduce the number of elements that represent the image (from thousands of pixels to a hundred of regions). The use of regions improves the robustness of the estimation of more complex features that will be used, first, when building the hierarchy and, afterwards, when analyzing the image. Furthermore, the boundaries of all relevant objects in the scene should be present in the initial partition, so that objects will be represented in the hierarchical structure by unions of these initial partition regions. Such a set of contours should be obtained with a small number of regions. Therefore, an adequate similarity measure has to be defined to move from the pixel level to a region-based representation that is coherent with the image content. Moreover, an automatic stopping criterion has to be proposed to avoid, as much as possible, oversegmentation and undersegmentation effects.

Selection of the merging criteria to build the hierarchy: The hierarchy of regions has to contain representations of the most relevant objects in the scene. This idea is linked with the concept of semantic analysis which, nowadays, is not a feasible task in unconstrained scenarios. Thus, in the context of generic image representation, merging criteria should be proposed that combine features being shared by the largest possible amount of semantic objects. Such a hierarchical representation would be afterwards used as starting structure for different specific image analysis procedures (e.g.: object detection).

After this introduction, Section 2 studies the most common criteria for the initial merging in a segmentation and proposes a new criterion, introducing an automatic technique for defining, given a merging criterion, its associated stopping criterion. Section 3 deals with the merging criteria that can be used to obtain a hierarchical representation of generic images. Each proposal is exemplified with particular cases as well as compared with other existing approaches. Finally, Section 4 drives some conclusions.

2. CREATION OF THE INITIAL PARTITION

In our work, the region model M_R is assumed to be constant within the region, and is the vector formed by the average values of all pixels $p \in R$, in the YCbCr color space.

This work has been partly supported by the EU project NoE MUSCLE FP6-507752 and by the grant TEC2004-01914 of the Spanish Government.

	Figure 2.a		Figure 2.b		Figure 2.c		Figure 2.d		Figure 2.e	
	$PSNR$	σ^2_{reg}	$PSNR$	σ^2_{reg}	$PSNR$	σ^2_{reg}	$PSNR$	σ^2_{reg}	$PSNR$	σ^2_{reg}
MSE	$21,79$	$0,757$	$22,71$	$2,959$	$24,86$	$6,212$	$21,44$	$0,531$	$28,15$	$1,753$
SE	$21,83$	$0,134$	$24,07$	$0,356$	$29,39$	$0,814$	$21,58$	$0,079$	$32,43$	$0,343$
RSE	$22,30$	$0,302$	$24,50$	$0,635$	$30,01$	$0,969$	$22,18$	$0,199$	$32,91$	$0,231$
L2	$22,06$	$0,098$	$24,06$	$0,221$	$29,50$	$0,442$	$22,13$	$0,115$	$32,29$	$0,085$

Table 1: Merging criteria comparison on the images from Figure 3. PSNR values are given in dBs. σ^2_{reg} values are divided by 10^6.

2.1. Initial Partition: Similarity measure

The similarity measure is computed for each pair of neighboring regions according to a selected homogeneity criterion. The basic criterion used in most segmentation approaches is color homogeneity. Some of the measures are size independent, like the mean squared error (MSE) between the merged region and its model [3] or the Euclidean distance between the region models [4]. In general, size independent color-based measures tend to produce partitions with few large regions and a large number of extremely small regions.

Trying to avoid this problem, other measures take into account the region sizes, as the squared error (SE) [3], or the weighted Euclidean distance [5]. When using these size dependent criteria, the cost of merging for small regions decreases, forcing small regions merge together first and encouraging the creation of large regions.

As a compromise between the two groups of measures, we propose a variant of the relative SE measure (RSE) [3], which shows a good balance between contour accuracy and size of final regions. In the sequel, this proposal is referred to as the L2 measure:

$$O(R_1,R_2)=N_{R_1}\|M_{R_1}-M_{R_1\cup R_2}\|_2+N_{R_2}\|M_{R_2}-M_{R_1\cup R_2}\|_2 \quad (1)$$

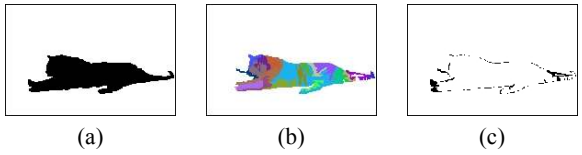
where N_{R_i} is the number of pixels in region R_i.

In the example of Table 1, different merging criteria are compared on a set of 100 images from the COREL database. To decouple the effects of the merging and stopping criteria, a simple stopping criterion is used (merge up to 50 regions). The comparison is performed in terms of final PSNR and of variance of the region sizes.

In terms of PSNR the L2 criterion outperforms the MSE criterion, obtains slightly better values than the SE criterion and slightly worst ones than the RSE criterion. However, these results are obtained while largely outperforming the three criteria in terms of variance of the region sizes. As previously commented, this is a relevant feature since ensures that regions obtained with the L2 measure will be adequate for a subsequent robust feature estimation: large enough and homogenous in size while presenting similar PSNR values than previous measures and leading to visually good representation (see results in the paper and in the web page http://gps-tsc.upc.es/imatge/_Veronica/ICIP2007.html). This behavior is illustrated in Figure 3 with 5 images of different complexities.

In the next experiment, the quality of the resulting partitions is assessed in terms of how accurately they represent semantic objects. We use the previous COREL database subset whose semantic objects (160 in total) have been manually segmented in the context of the SCHEMA project (http://www.iti.gr/SCHEMA/). The set contains 10 images of 10 different complexity classes which are grouped and ordered in the following way: *tigers*, *horses*, *eagles*, *mountains*, *fields*, *cars*, *jets*, *beaches*, *butterflies* and *roses*.

To asses regions conformance to the semantic object boundaries, the distance proposed in [6] is used. It proposes a symmetric distance for comparing partitions which is extended to an asymmetric distance in a way that the distance between one partition and any partition finer than it is zero.Thus, this distance is appropriate for our purposes since it provides with a coherent framework for comparing both the initial partitions (in Section 2) and the selected BPT nodes (in Section 3) with the object partitions manually obtained. An example of how this distance is computed is presented in Figure 1. The two partitions are compared in terms of the amount of pixels that should change their labels to have a perfect contour match between partitions. The final distance is normalized by the object size.



(a) (b) (c)

Fig. 1: Example of asymmetric distance: (a) Object partition. (b) Regions from the initial partition matching the object partition. (c) Pixels requiring a label change.

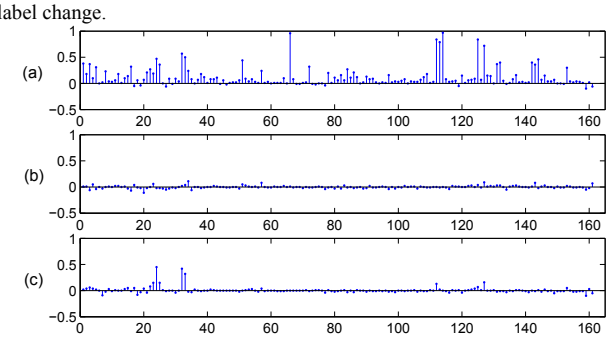


Fig. 2: Difference between the asymmetric distance values. (a) MSE-L2. (b) SE-L2. (c) RSE-L2.

Figure 2 shows the difference between the asymmetric distance values obtained using the three previous merging criteria and the L2 merging criteria. Statistics of each merging criterion as well as of the differences are presented in Table 3. Note that, in this case, the global behavior of the SE and L2 criteria is very similar, outperforming those of MSE and RSE criteria.

	SE	MSE	RSE	L2	SE-L2	MSE-L2	RSE-L2
Mean	10.52	22.57	11.46	10.53	-0.01	12.04	0.97
σ^2	1.012	5.549	1.880	1.038	0.074	3.633	0.403

Table 2: Merging criteria comparison on the COREL subset. Asymmetric distance mean and variance values are multiplied by 10^2.

2.2. Initial Partition: Stopping Criterion

Typical stopping criteria deal with reaching an a priori value of a parameter such as the final number of regions or the global PSNR. However, as we are creating an initial partition in the hierarchical representation, the objective is to segment the image into regions corresponding to parts of the objects in the scene whilst avoiding the creation or regions spanning more than one object. Thus, the stopping criterion has to take into account the complexity of the scene.

We propose a procedure to estimate this complexity based on the *accumulated merging cost*. Let $O(k)$ be the cost of the merging at

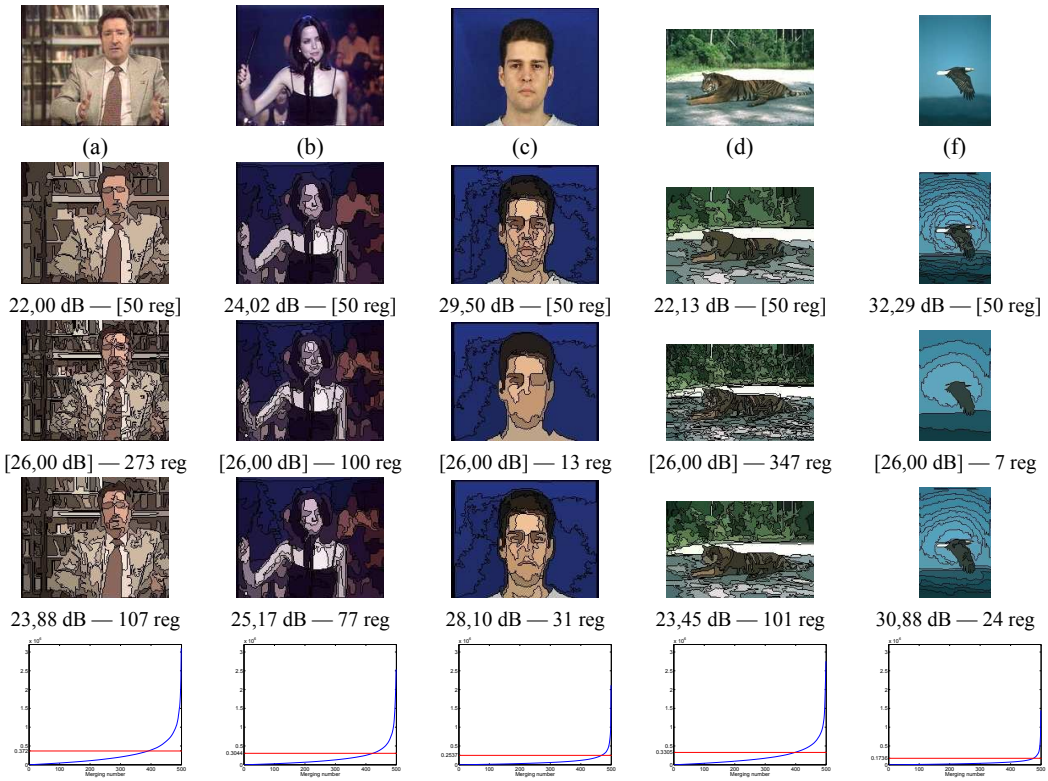


Fig. 3: Comparison of the different stopping criteria. First row: original images. Second row: SC1: Nreg = 50. Third row: SC2: PSNR = 26 dB. Fourth row: SC3: T_{ACM} = 0.12. Fifth row: AMC(m).

iteration k. The accumulated merging cost (AMC) is defined as

$$AMC(m) = \sum_{k=1}^{m} O(k). \quad (2)$$

The stopping criterion is defined as a fraction $T_{AMC} \in [0, 1]$ of the total AMC ($AMC(N-1)$), where N is the image size. This criterion stops the merging process at iteration $\bar{m}$, where

$$\bar{m} = min\{m/AMC(m) > AMC(N-1)T_{AMC}\}. \quad (3)$$

Note that if the similarity measure used in the merging process is the relative squared error (RSE) [3], then the accumulated cost equals the squared error and the stopping criteria becomes a threshold relative to the maximum PSNR. A stopping criterion based on the analysis of the accumulated cost was also proposed in [5]. However, this approach is not useful in our case since the resulting partitions in [5] have a very reduced number of regions.

Although the exact computation of the new criterion requires computing and storing the whole sequence of fusions and merging costs, note that the initial fusions are directly related to merging pixels and have a very low merging cost. Thus, we can obtain first a fine partition (with small, homogeneous regions) and then find the initial partition by merging regions from this fine partition, only storing information of these last mergings.

Figure 3 compares, for a set of images with different complexity, the results obtained by the most common stopping criteria (SC1: final number of regions Nreg and SC2: final PSNR) and the proposed criterion (SC3). As it can be seen, the proposed criterion adapts to the image complexity (that is, it avoids oversegmentation as well as undersegmentation effects) obtaining partitions in which the main objects in the scene are correctly represented. The last row of Figure 3 shows the accumulated costs plotted for each iteration of the merging process, starting with a fine partition composed of 500 regions. Plots also show the thresholds obtained for $T_{AMC} = 0.12$. This value has been selected after analyzing the robustness of the system with respect to its variations, as presented in Figure 4.

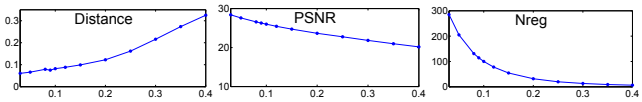


Fig. 4: Analysis of the T_{AMC} impact.

The proposed stopping criterion is assessed, as in Subsection 2.1, by using the COREL subset and the asymmetric distance. In this case, presented in Figure 5, the similarity measure is fixed (L2) whereas we compare as stopping criteria (a) a fixed Nreg and (b) a fixed PSNR with the proposed AMC. The Nreg and PSNR used values are the mean values obtained by the AMC criteria over the COREL database subset (Nreg = 77 and PSNR = 25.54 dB).

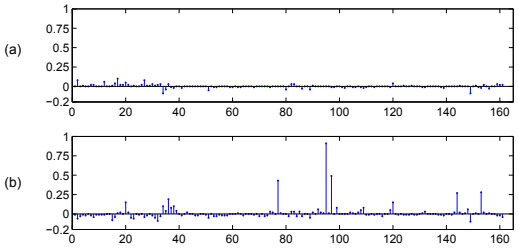


Fig. 5: Difference between the asymmetric distance values. (a) Nreg-AMC. (b) PSNR-AMC.

Plots in Figure 5 show that, for complex (simple) images, the AMC criterion outperforms the Nreg (PSNR) criterion. This is the

	Nreg	PSNR	AMC	Nreg-AMC	PSNR-AMC
Mean	8.89	10.37	8.81	0.11	1.55
σ^2	0.66	1.70	0.57	0.04	0.98

Table 3: Stopping criteria comparison on the COREL subset. Asymmetric distance mean and variance values are multiplied by 10^2.

case of the classes *tigers* and *horses* (*eagles* and *jets*) where the amount of regions (PSNR) is too low producing undersegmented results that lead to higher asymmetric distances (see examples in Figure 3). In turn, Table 3 shows that in general the AMC criterion outperforms the PSNR criterion, while slightly improving the Nreg criterion.

3. BUILDING THE HIERARCHICAL REPRESENTATION

Here, we analyze the construction of the hierarchical structure from the regions defined by the initial partition. The discussion is centered on the similarity measure. Ideally, nodes in the tree should be objects or parts of objects with a semantic meaning. Therefore, the similarity measure should be related to a notion of object. Several approaches to segmentation try to create 'meaningful' partitions incorporating geometric features into the segmentation process, like measures of proximity, compactness, inclusion or symmetry (see [5]). However, the integration of this information is difficult to analyze and evaluate, since there is a strong overlap between various geometrical features (adjacency, contour complexity, quasi-inclusion).

The proposed merging criterion has two terms. One is based on color similarity. The color difference in each component is normalized by the dynamic range of the component in the image. This way, it adapts to the chrominance variability of the image. For each image component we compute the L^2 norm between each region and its model normalized by the component dynamic range.

$$\begin{aligned} O_{color}(R_1,R_2) &= N_{R_1}\|\mathbf{w}(M_{R_1}-M_{R_1\cup R_2})\|_2 \\ &+ N_{R_2}\|\mathbf{w}(M_{R_2}-M_{R_1\cup R_2})\|_2 \end{aligned} \quad (4)$$

$\mathbf{w}$ is a weight vector where w_i is the inverse of the dynamic range of the image component $i \in \{Y, Cb, Cr\}$.

The second term is related to the contour complexity of the merged regions. After analyzing several approaches, a measure has been adopted that computes the increase in perimeter of the new region with respect to the largest of the two merged regions: $\Delta P(R_1,R_2) = min(P_1,P_2) - 2P_{12}$, where P_1 and P_2 are the R_1 and R_2 perimeters, respectively, and P_{12} is the common perimeter between the regions. The term that measures contour complexity is

$$O_{cont}(R_1,R_2) = max(0, \Delta P(R_1,R_2)) \quad (5)$$

which sets to 0 negative increments that occur when a region is partially or totally included in the other.

Color and contour similarity measures are linearly combined by:

$$O(R_1,R_2) = \alpha O_{color}(R_1,R_2) + (1-\alpha)O_{cont}(R_1,R_2) \quad (6)$$

where $\alpha \in [0,1]$, and typically $\alpha = 0.5$ is used.

Figures 6 and 7 present an example of BPT created with the new and the L2 criteria, respectively. It exemplifies the case of objects in the scene being correctly gathered in single nodes whereas, using the L2 criterion, their information was split among various nodes. The analysis of these criteria on the COREL subset is presented in Table 4. For each image in the database, the node in the BPT leading to the smallest symmetric distance has been selected. Note that the new criterion outperforms both the RSE and the L2 criteria.

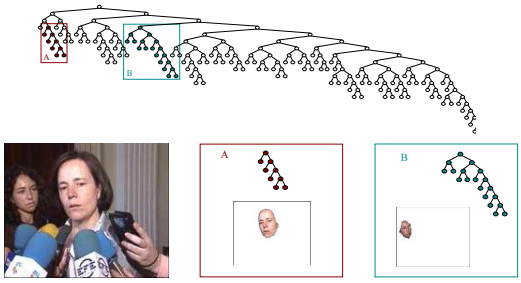


Fig. 6: BPT created with the new similarity measure of equation 3.

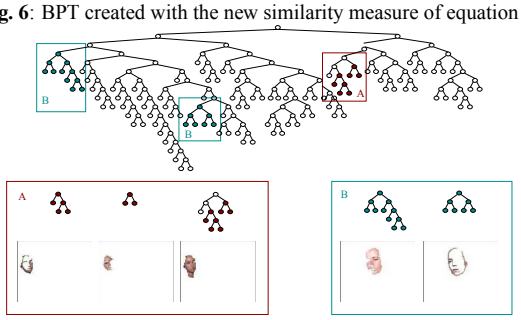


Fig. 7: BPT built with the L2 similarity measure.

	RSE	L2	New	RSE-New	L2-New
Mean	24.62	25.65	20.40	4.22	5.25
σ^2	4.45	4.30	3.20	2.15	2.23

Table 4: Symmetric distance over the BPT on the COREL subset. Mean and variance values are multiplied by 10^2.

4. CONCLUSIONS

This paper has analyzed the creation of a region-based generic-image hierarchical representation. The combination of the proposed merging and stopping criteria for the initial partition and the merging criterion for the BPT creation produce hierarchical descriptions useful for object detection purposes. This hierarchical representation has already been successfully used in a face detection algorithm. Currently, we are extending it to a generic object detection algorithm in which, although objects may not be completely represented in a BPT node, the best node in the BPT is used as an initial estimate of the location and size of the object.

5. REFERENCES

[1] P. Salembier and F. Marqués, "Region-based representation of image and video: Segmentation tools for multimedia services," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 8, pp. 1147–1167, December 1999.

[2] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation and information retrieval," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 561–575, April 2000.

[3] L. Garrido, *Hierarchical Region Based Processing of Images and Video Sequences: Application to Filtering, Segmentation and Information Retrieval*, Ph.D. thesis, UPC, April 2002.

[4] T. Vlachos and A. G. Constantinides, "Graph-theoretic approach to color picture segmentation and contour classification," in *IEE Proceedings*, February 1993, vol. 130, pp. 36–45.

[5] T. Adamek, *Using Contour Information and Segmentation for Object Registration, Modeling and Retrieval*, Ph.D. thesis, Dublin City University, June 2006.

[6] J. Cardoso and L. Corte-Real, "Toward a generic evaluation of image segmentation," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1773 – 1782, November 2005.