

Hierarchical Image Segmentation Relying on a Likelihood Ratio Test

Silvio Jamil F. Guimarães^{1,2}(✉), Zenilton Kleber G. do Patrocínio Jr.¹, Yukiko Kenmochi², Jean Cousty², and Laurent Najman²

¹ PUC Minas - ICEI - DCC - VIPLAB, Belo Horizonte, Brazil
{sjamil,zenilton}@pucminas.br

² Université Paris-Est, LIGM, ESIEE Paris - CNRS, Champs-sur-Marne, France
{y.kenmochi,j.cousty,l.najman}@esiee.fr

Abstract. Hierarchical image segmentation provides a set of image segmentations at different detail levels in which coarser details levels can be produced by simple merges of regions from segmentations at finer detail levels. However, many image segmentation algorithms relying on similarity measures lead to no hierarchy. One of interesting similarity measures is a likelihood ratio, in which each region is modelled by a Gaussian distribution to approximate the cue distributions. In this work, we propose a hierarchical graph-based image segmentation inspired by this likelihood ratio test. Furthermore, we study how the inclusion of hierarchical property have influenced the computation of quality measures in the original method. Quantitative and qualitative assessments of the method on three well known image databases show efficiency.

Keywords: Hierarchical image segmentation · Graph-based method · Statistical properties

1 Introduction

Image segmentation is the process of grouping perceptually similar pixels into regions. A hierarchical image segmentation is a set of image segmentations at different detail levels in which the segmentations at coarser detail levels can be produced from simple merges of regions from segmentations at finer detail levels. Therefore, the segmentations at finer levels are nested with respect to those at coarser levels. Hierarchical methods have the interesting property of preserving spatial and neighboring information among segmented regions. Here, we propose a hierarchical image segmentation in the framework of vertex-edge-weighted graphs, where the image is equipped with an adjacency graph, the cost of an edge is given by a dissimilarity between two points of the image and the cost of a vertex is the color information of the associated point. Therefore, the adjacency graph is represented by data structures in order to efficiently compute this hierarchy.

The authors are grateful to PUC Minas, CNPQ, CAPES and FAPEMIG for the partial financial support of this work.

The first appearance of minimum spanning tree in pattern recognition for representing a hierarchy dates back to the seminal work of Zahn [19]. Lately, its use for image segmentation was introduced by Morris *et al.* [14] in 1986 and popularized in 2004 by Felzenszwalb and Huttenlocher [9], Noch and Nielsen [16] proposed a statistical method in which the merging order is similar to the creation of a MST. However the region-merging method [9, 16] does not provide a hierarchy. Some optimality properties of hierarchical segmentations have been studied in [6, 15]. New characterizations between MST and saliency maps based on quasi-flat zones have been studied in [8]. Considering that, for a given image, one can tune the parameters of the well-known method [9] for obtaining a reasonable segmentation of this image. A seminal framework to transform a non-hierarchical method to a hierarchical one has been proposed in [11]. Following a similar idea, we proposed in [12] a method for hierarchizing the approach proposed in [16] in which the image segmentation is formulated as an inference problem. In [5, 17] were proposed methods, that can also be formulated as inference problems, relying on likelihood ratio test. In both cases, the regions are iteratively merged until a termination criterion is fulfilled. Unlike the method proposed in [5], a one proposed in [17] does not directly use the likelihood ratio test as similarity measure taking it as an enhancement for the merging evidence. Thus, the proposed predicate can be interpreted as a combination of consistency and similarity measures since these ones are computed from pixel values randomly sub-sampled in each pair of tested regions. The existent consistency tells whether the tested data belong to the same group, and it is measured by two hypotheses according to the sequential probability ratio test. Moreover, a Gaussian distribution model to approximate the cue distributions has been used in [17] instead of Kullback-Leibler divergence, which is used in [5]. Furthermore, the method proposed in [17] holds certain global properties, *i.e.*, by using the merging predicate the results are neither overmerged nor undermerged, which preserves the perceptual cues.

Even if the results presented in [5, 17] are interesting since they have important statistical properties, the merging order is adaptive since, after each merging step, the similarity measure between all adjacent regions must be updated in order to identify the two new adjacent regions of maximum similarity value. Thus, in this work, instead of considering an adaptive merging order, we propose a hierarchical method in which the merging order (or order for evaluating the regions) is defined, a priori, by the weights of MST computed from the image. Unfortunately, the causality and the location properties are missing in [17] whether their two parameters (λ_1 and λ_2) are considered as scales, as can be seen in Fig. 1. According to [17], the number of regions decreases when λ_1 increases, and the number of regions increases when λ_2 decreases, however these statements are not completely true as showed in Fig. 1. According to [10], a “scale” is considered as a true scale-parameter, when it satisfies both the causality principle and the location principle, which leads to work with a hierarchy of segmentations. In this sense, the method proposed in [17], so-called **SPRT**, does not produce a hierarchy of partitions.

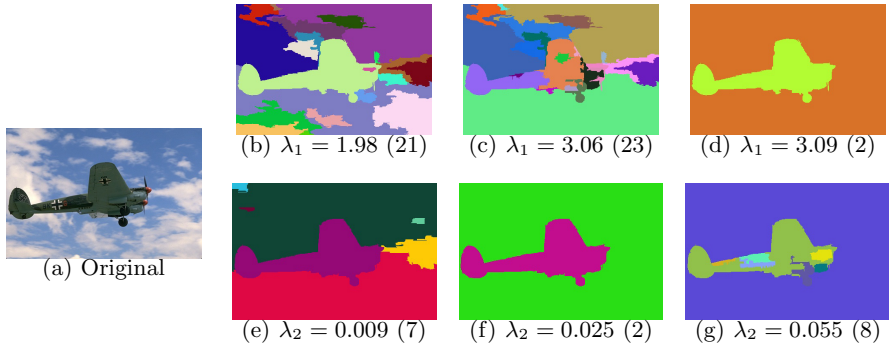


Fig. 1. A real example illustrating the violation of the causality and location principles by [17] in which the number of regions (in parentheses) is not monotonic, when the so-called “segmentation scale” increases. Moreover, the location of the contours are not stable in different segmentations. In first row, the parameter λ_2 is equal to 1 and in the second row, the parameter λ_1 is equal to 1.

The main contribution of this paper is the proposal of a similarity measure based on likelihood ratio test computed from a Gaussian model distribution in the context of graph-based hierarchical segmentation. Moreover, unlikely [5, 17], instead of iteratively deciding whether two adjacent regions might be merged, we compute the scales for which the regions must be merged. Differently of [17], our method can be directly applied to the original image instead of computing on an over-segmentation. Although, our method provides all statistical scales is more efficient than **SPRT** and according to our experiments the proposed method is statistically better, in terms of paired t-test analysis, than **SPRT**.

Furthermore, since our algorithm is a hierarchical approach, its result satisfies both the locality principle and the causality principle. Namely, the number of regions decreases when the scale parameter increases, and the contours do not move from one scale to another, as can be seen in Fig. 2.

This work is organized as follows. In Section 2, we present our hierarchical method for color image segmentation. Some experimental results performed on three well known image databases are given in Section 3. Finally, in Section 4, some conclusions are drawn and further works are discussed.

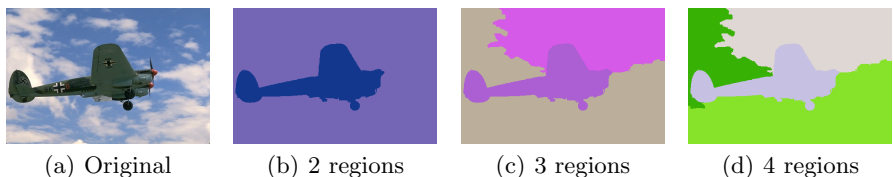


Fig. 2. A real example illustrating the hierarchical segmentation obtained by our proposed method inspired by similarity measure proposed in [17] showing the number of regions. Moreover, the location of the contours are stable in different segmentations.

2 A Method Inspired by Likelihood Ratio Test

In this section we present the (dis)similarity measure which is used to verify if two regions must be merged. We also propose a method that computes the low scale in which two regions must be merged without violating the likelihood test between these two regions. Moreover, we compute a hierarchy of partitions providing all scales instead of only one segmentation.

2.1 Likelihood Ratio Test

The likelihood ratio test computed from two regions in which the Gaussian distribution model is used to approximate the cue distributions x has been proposed in [17], as follows

$$\begin{aligned} P_0(x|\theta_0) &= \lambda_1 \exp(-(I_Y - I_{X \cup Y})^T S_I^{-1} (I_Y - I_{X \cup Y})) \\ P_1(x|\theta_1) &= 1 - \lambda_2 \exp((I_Y - I_X)^T S_I^{-1} (I_Y - I_X)) \end{aligned} \quad (1)$$

in which I_X and I_Y are the average color of the samples in regions X and Y , respectively, and $I_{X \cup Y}$ is the average value of the samples' union. θ is called hypothesis in which θ_0 is related to consistent regions and θ_1 is related to inconsistent regions. S_I is the covariance matrix of the regions, λ_1 and λ_2 are scalar parameters which could be considered as “scales” in the segmentation process. If each test is independent, the composition of the likelihood ratio δ is the sum of the individual $\delta_i = \log \frac{P_0(x_i|\theta_0)}{P_1(x_i|\theta_1)}$ for N iterations, in which N is the first integer for which $\delta \geq A$ (consistent regions) or $\delta \leq B$ (inconsistent regions). It is possible to see that the solution to the hypothesis is decided by the relationship between δ and a pair of upper and lower limits denoted by A and B , respectively. If δ goes out of one of these limits, the hypothesis is made, and thus, the test stops. Otherwise, the test is carried on with a new random sampling. Due to space limitations, more explanations are omitted here (see [17] for more details).

2.2 Similarity Measure

Before discussing about the similarity measure, let's remember some definitions. Let $A = \log \frac{1-\beta}{\alpha}$ be one value computed from two constants α and β , which are probabilities of the decision error and are set to 0.05 [17]. Let X and Y be two neighboring regions. Let $P_0(x|\theta_0)$ and $P_1(x|\theta_1)$ be two conditional probabilities for representing the Gaussian distribution model to approximate the cue distributions x . From Eq. 1, $P_0(x|\theta_0) = \lambda_1 \exp(-D_Y^{X \cup Y})$ in which $D_Y^{X \cup Y} = (I_Y - I_{X \cup Y})^T S_I^{-1} (I_Y - I_{X \cup Y})$ could be computed by the difference between color averages of the regions Y and $X \cup Y$, and it represents the consistency of region X with respect to Y . Analogously, $P_1(x|\theta_1) = 1 - \lambda_2 \exp(-D_Y^X)$ in which $D_Y^X = (I_Y - I_X)^T S_I^{-1} (I_Y - I_X)$. Considering that the parameters λ_1 and λ_2 control the coarseness (or “scale”) of a segmentation, it is important to understand the influence of these parameters in the process. Moreover, we

consider that the number of iterations, N , is equal to 1. Thus, without loss of generality, we fix λ_2 , and the distribution of visual cues for producing consistent regions ($\delta \geq A$) could explicitly define the parameter λ_1 , as follows:

$$\delta \geq A \tag{2}$$

$$\frac{\lambda_1 \exp(-D_Y^{X \cup Y})}{1 - \lambda_2 \exp(-D_Y^X)} \geq \frac{1 - \beta}{\alpha} \tag{3}$$

$$\lambda_1 \geq \frac{1 - \beta}{\alpha} \times \frac{1 - \lambda_2 \exp(-D_Y^X)}{\exp(-D_Y^{X \cup Y})} \tag{4}$$

Fig. 1 illustrates the missing of the location and the causality principles, for variation of both λ_1 and λ_2 , which controls the so-called segmentation scale, and this shows the absence of hierarchical properties of the method **SPRT**. In fact, this work will look for a procedure to adapt the values of λ_1 according to the analyzed regions in order to guarantee that two regions are correctly merged. Then, the scale $Q_Y(X)$ of X relative to Y , which represents λ_1 when λ_2 is equal to 1, is defined as:

$$Q_Y(X) = \frac{1 - \beta}{\alpha} \times \frac{1 - \exp(-D_Y^X)}{\exp(-D_Y^{X \cup Y})} \tag{5}$$

Thus, the scale $Q(X, Y)$ for merging two regions X and Y could be written by

$$Q(X, Y) = \max(Q_Y(X), Q_X(Y)) \tag{6}$$

2.3 The Proposed Method

In this section we describe our method, so-called **hPRT**, to compute a hierarchy of partitions based on scales, so-called here *probability ratio scale*, as defined by Eq. 6. The main difference between this method and our previous work [12] is the application of a new measure based on likelihood probability ratio test. Let us first recall some important notions for handling hierarchies [4, 6, 7, 14, 15].

According to [8], for any tree T spanning the set V of the image pixels, to any map $w : E \rightarrow \mathbb{N}$ that relates a weight to each edge of T , one may associate the partition \mathbf{P}_λ^w of V for a given threshold $\lambda \in \mathbb{N}$, induced by the connected components of the graph made from V and edges whose weights are below λ . It is well known [6, 14] that for any two values λ_1 and λ_2 such that $\lambda_1 \geq \lambda_2$, the partitions $\mathbf{P}_{\lambda_1}^w$ and $\mathbf{P}_{\lambda_2}^w$ are *nested* and $\mathbf{P}_{\lambda_1}^w$ is *coarser* than $\mathbf{P}_{\lambda_2}^w$. Hence, the set $\mathcal{H}^w = \{\mathbf{P}_\lambda^w \mid \lambda \in \mathbb{N}\}$ is a *hierarchy of partitions induced by the weight map w* . Each element of a partition is called a *region*, and the *index* of a region A is the largest weight of the edges of the subtree induced by A from T . We also denote by $R_{\mathcal{H}^w}(x, \lambda)$ the set of all regions of \mathcal{H}^w which contains x and whose index is less than λ . Our algorithm does not explicitly produce a hierarchy of partitions, instead it produces a new weight map L (scales of probability ratio values) from which the desired hierarchy \mathcal{H}^L can be inferred. It starts from a

minimum spanning tree T of the gradient edge-weighted graph built from the image. In order to compute the scale $L(e)$ associated with each edge of T , after an initialization step, our method iteratively considers the edges of T in a non-decreasing order of their original weights w . For every edge e , the new weight $L(e)$ of e is initialized to ∞ ; then, for each edge e linking two vertices x and y the following steps are proceeded:

1. Compute the probability ratio scale s of e with respect to the current values of L :

$$s = \min\{\lambda \mid C(x, y, \lambda, L) \text{ is true}\} \quad (7)$$

where $C(x, y, \lambda, L)$ is true if

$$\forall X \in R_{\mathcal{H}^L}(x, \lambda), \forall Y \in R_{\mathcal{H}^L}(y, \lambda), Q(X, Y) \leq \lambda.$$

2. Update the weight of e for L with the value s obtained at step 1:

$$L(e) = s.$$

Intuitively, the probability ratio scale at edge e corresponds to the lowest scale value λ such that we cannot find two regions in the current hierarchy (the one associate to L at the current iteration) which are linked by e and whose similarity is greater than λ (according to the similarity measure Q). Hence, the hierarchical probability ratio scale $L(e)$ is computed based on a minimization procedure related to the similarity measure Q on the set of all the possible pairs of regions linked by the edge e .

3 Experimental Analysis

In this section, we have done assessments in terms of F-measure and precision-recall curves by using six segmentation methods applied to three different image databases.

3.1 Compared Methods and Their Underlying Graphs

As mentioned before, we will compare our method **hPRT** that provides a hierarchical graph-based segmentation result with the original method **SPRT** [17] that provides a non-hierarchical graph-based segmentation result, even if in the original paper the method is presented as hierarchical, we have show that two features of hierarchical methods are missing, the location and the causality principles (see Fig. 1). We also compare the proposed method to **GB** [9], **hGB** [11], **SRG** [17] and **hSRG** [12]. The parameters of each non-hierarchical method are given below:

GB. This method depends on three parameters: (i) k ; (ii) σ ; and (iii) τ . The parameter k , which control the the so-called “scale” varies in [100, 10000]. Gaussian smoothing pre-processing with parameter σ , whose values are set to 0 (no smoothing). The area parameter τ , which is detailed ahead, is used to merge adjacent regions.

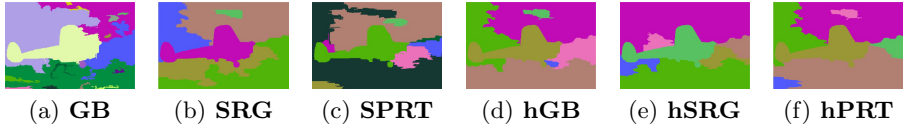


Fig. 3. Some results by using the compared segmentation methods (**GB**, **hGB**, **SRG**, **hSRG**, **SPRT** and **hPRT**). In order to compute these results, we consider the parameters that obtain the best F-measures for each method. Moreover, we illustrate the segmentations containing either 7 regions (easy for hierarchical methods) or as close as possible to 7 regions.

SRG. This method depends basically on the parameter q , whose values vary in $[2, 256]$.

SPRT. This method depends on two parameters, λ_1 and λ_2 . In order to study the behaviour of these parameters, we tested two different configurations. In the first one, we fixed the parameter $\lambda_1 = 1$ and λ_2 varies in $[0.001, 5]$. In the second one, we fixed the parameter $\lambda_2 = 1$ and λ_1 varies in $[0.1, 10]$. However, to apply this method, a pre-processing is needed for produce a over-segmented image, in order to do that, we apply **GB** in which $k = 100$ and σ vary in $[0.1, 1]$.

For the hierarchical methods, we consider the most simple configuration by ignoring, for example, the smoothing pre-processing. In all cases, an area-filtering post-processing step with parameter τ , which is the ratio of the component size to the image size, is applied. The values of τ are set to 0.01%, 0.05% or 0.1% of image size. This post-processing is an important step since the order to analyse the edges is pre-defined and depends on the gradient values, thus small regions which contain elements with high gradient values could have high values of scales.

Before applying those methods, it is necessary to transform a given image into an edge-weighted graph for **hGB** and **GB** or into a vertex-edge-weight graph for **SRG** and **hSRG**. In this paper, we consider the following underlying graph. The graph is induced by the 8-adjacent pixel relationship, where each vertex corresponds to a pixel and each edge corresponds to a pair of adjacent pixels. Each edge is weighted by a simple color gradient: the Euclidean distance in the RGB space between the colors of the two adjacent pixels.

In order to illustrate some segmentations, Fig. 3 shows some results obtained by the compared methods. Due to the features of those methods, it was not possible to obtain exactly 7 regions, thus we present segmentations containing as close as possible the number of required regions. Fig. 4 illustrates some results using only the hierarchical methods. As can be seen, the two yellow flowers (center and right of the image) are better identified when **hPRT** is used since it is necessary only 4 segments (including the background) against 6 and 9 segments for **hSRG** and **hGB**, respectively. In other words, for this example, the proposed method is more robust than the others to obtain the single objects.

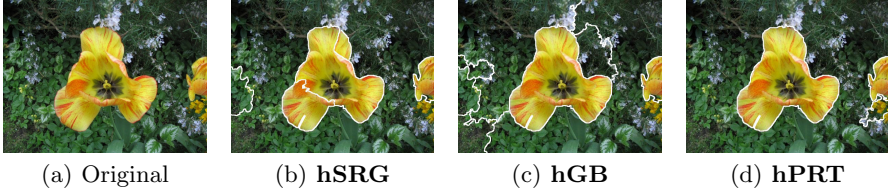


Fig. 4. Some results by using the hierarchical compared segmentation methods (**hGB**, **hSRG** and **hPRT**). In order to compute these results, we consider the parameters that makes possible to obtain the yellow flowers (center and right of the original image).

Table 1. Performances of our method **hPRT** and the compared methods (**GB**, **hGB**, **SRG**, **hSRG** and **SPRT**) using Ground-truth Covering (GT Covering). The presented scores in (a) are optimal considering a constant scale parameter for the whole dataset (ODS) and a scale parameter varying for each image (OIS). In (b), the best method choice is presented with confidence interval for each pairwise comparison. See [3] for more details on the evaluation method.

(a)					(b)				
Database	Method	GT Covering		F-Measure	Database	Methods	F-measure for regions		
		ODS	OIS				Best	Confidence interval (5%)	The best method
BSDS500	GB	0.42	0.54	0.68	0.59	BSDS500	GB x hPRT	[-0.0213, 0.00519]	equivalent
	SRG	0.51	0.57	0.68	0.63		SRG x hPRT	[0.0249, 0.0486]	SRG
	SPRT	0.45	0.52	0.57	0.57		SPRT x hPRT	[-0.0397, -0.0163]	hPRT
	hGB	0.45	0.54	0.63	0.60		hGB x hPRT	[0.000872, 0.0171]	hGB
	hSRG	0.45	0.53	0.61	0.59		hSRG x hPRT	[-0.0169, 0.000151]	equivalent
	hPRT	0.45	0.53	0.61	0.60				
GRABCUT	GB	0.72	0.77	0.79	0.79	GRABCUT	GB x hPRT	[-0.0307, 0.019]	equivalent
	SRG	0.71	0.75	0.77	0.74		SRG x hPRT	[-0.115, -0.0145]	hPRT
	SPRT	0.72	0.76	0.78	0.77		SPRT x hPRT	[-0.0582, 0.00445]	equivalent
	hGB	0.71	0.77	0.81	0.79		hGB x hPRT	[-0.0231, 0.0165]	equivalent
	hSRG	0.71	0.74	0.77	0.75		hSRG x hPRT	[-0.078, -0.0147]	hPRT
	hPRT	0.73	0.78	0.80	0.80				
WI1OBJ	GB	0.67	0.75	0.78	0.75	WI1OBJ	GB x hPRT	[-0.0187, 0.0483]	equivalent
	SRG	0.68	0.73	0.75	0.73		SRG x hPRT	[-0.0286, 0.0237]	equivalent
	SPRT	0.63	0.68	0.72	0.70		SPRT x hPRT	[-0.0534, -0.0112]	hPRT
	hGB	0.64	0.72	0.76	0.74		hGB x hPRT	[-0.0122, 0.0192]	equivalent
	hSRG	0.64	0.70	0.73	0.71		hSRG x hPRT	[-0.035, -0.00269]	hPRT
	hPRT	0.65	0.71	0.74	0.73				
WI2OBJ	GB	0.74	0.85	0.88	0.85	WI2OBJ	GB x hPRT	[-0.0283, 0.0194]	equivalent
	SRG	0.78	0.84	0.86	0.84		SRG x hPRT	[-0.0396, 0.00904]	equivalent
	SPRT	0.73	0.82	0.84	0.82		SPRT x hPRT	[-0.0517, -0.0149]	hPRT
	hGB	0.74	0.86	0.88	0.86		hGB x hPRT	[-0.015, 0.0219]	equivalent
	hSRG	0.77	0.83	0.86	0.84		hSRG x hPRT	[-0.0407, 0.00103]	equivalent
	hPRT	0.76	0.85	0.87	0.86				

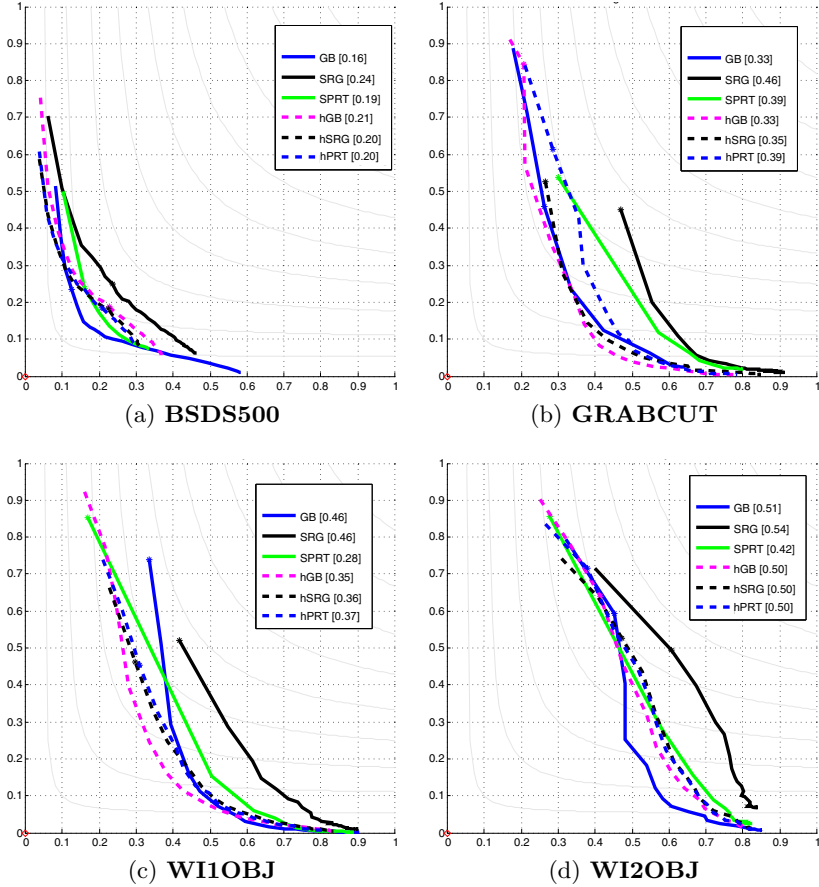


Fig. 5. Precision-Recall curves for objects and parts computed on three databases. The curves represent the 5 (five) segmentation methods (**GB**, **hGB**, **SRG**, **hSRG** and **SPRT**) and the proposed method (**hPRT**). The marker on each curve is placed on the Optimal Dataset Scale (ODS). Moreover, the F-measures of the marked points on each curve is presented in brackets.

3.2 Databases

In order to provide a comparative analysis between several methods, we used three different databases: (i) the Berkeley Segmentation Dataset [13], called **BSDS500**; (ii) the database proposed in [18], called **GRABCUT**; and (iii) the database proposed in [1, 2] which is divided into two groups – single and two objects – called **WI1OBJ** and **WI2OBJ**, respectively.

3.3 Quantitative Analysis

In this section, we assess the proposed method **hPRT** when compared to some other methods by using Ground-truth Covering (GT Covering) as showed in

Table 1. According to [3], the GT Covering of a segmentation S by a family of ground-truth segmentations $\{G_i\}$ is defined by first covering S separately with each human segmentation G_i , and then averaging over the different humans (see [3] for more details). Furthermore, we present the best method choice with confidence interval for each pairwise comparison. By using the F-measure, which is the harmonic mean of precision and recall, the compared methods are quite similar to the proposed one (as can be observed in Table 1(a)), however according to the pairwise comparison, the proposed method **hPRT** is always better than **SPRT**, and similar to other methods for all databases, except for **BSDS500** (as can be observed in Table 1(b)). Usually, the F-measure can be seen as a summary statistic of each method.

When the comparison is made by using the precision-recall curves for object and parts (Fig. 5), the method **hPRT** is better than or equivalent to **SPRT**. For other methods, test results are inconclusive for all databases.

In order to illustrate an example of computation time, we implemented our algorithm in C++ on a standard single CPU computer under OS X, we run it in a Intel Core i5, 4GB. For the image shown in 1 (a) (with size 321x481), the hierarchy is computed in 5 seconds, without any pre-processing. Considering that the method proposed in [17] is highly dependent on the number of regions in the over-segmented image, we present three different measurements: (i) for 722 regions the time is 3 seconds; (ii) for 1204 regions the time is 7 seconds; and (iii) for 2765 regions the time is 24 seconds.

4 Conclusions and Further Works

In this work, we propose a method for transforming a non-hierarchical method into a hierarchical one preserving the merging criterium, *i.e.*, all regions are merged according to the same probability ratio criterium. Differently of the method that iteratively computes the hierarchies and following our previous works, our method produces a weight map L (scales of probability ratio values) from which the desired hierarchy can be easily inferred.

According to our results, the inclusion of the hierarchical property on this region merging approach solves the causality and the location problems which are missing in **SPRT** method, without prejudicing the quality of results, in fact, our method **hPRT** is statistically better than **SPRT** when F-measure is used for comparison, and equivalent for other compared methods in three databases and worst than **SRG** and **hGB** for **BSDS500**. Moreover, unlikely the original method, our hierarchical one is not dependent on the over-segmented image to produce the segmentation.

In all the tests performed in this paper, we filter out small regions at all levels of the hierarchies. This filtering step has a strong impact on the quality of the results, that deserves an in-depth study by itself: it is actually a transformation of the hierarchy, and as an operator acting on hierarchies, it has clearly some theoretical and practical properties. In future work, we endeavor doing such a study. Other items that should be the topic of further studies are the robustness to noise, as well as the choice of a good hierarchical scale.

References

1. Alpert, S., Galun, M., Basri, R., Brandt, A.: Image segmentation by probabilistic bottom-up aggregation and cue integration. In: CVPR, June 2007
2. Alpert, S., Galun, M., Brandt, A., Basri, R.: Image segmentation by probabilistic bottom-up aggregation and cue integration. *PAMI* **34**(2), 315–327 (2012)
3. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. *PAMI* **33**, 898–916 (2011)
4. Beucher, S.: Watershed, hierarchical segmentation and waterfall algorithm. In: Proceedings of the 2nd International Symposium on Mathematical Morphology and Its Applications to Image Processing, ISMM 1994, Fontainebleau, France, September 1994, pp. 69–76 (1994)
5. Calderero, F., Marques, F.: Region merging techniques using information theory statistical measures. *Trans. Img. Proc.* **19**(6), 1567–1586 (2010)
6. Cousty, J., Najman, L.: Incremental algorithm for hierarchical minimum spanning forests and saliency of watershed cuts. In: Soille, P., Pesaresi, M., Ouzounis, G.K. (eds.) ISMM 2011. LNCS, vol. 6671, pp. 272–283. Springer, Heidelberg (2011)
7. Cousty, J., Najman, L.: Morphological floodings and optimal cuts in hierarchies. In: ICIP, pp. 4462–4466 (2014)
8. Cousty, J., Najman, L., Kenmochi, Y., Guimarães, S.: New characterizations of minimum spanning trees and of saliency maps based on quasi-flat zones. In: Benediktsson, J.A., Chanussot, J., Najman, L., Talbot, H. (eds.) ISMM 2015. LNCS, vol. 9082, pp. 205–216. Springer, Heidelberg (2015)
9. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. *IJCV* **59**, 167–181 (2004)
10. Guigues, L., Cocqueruz, J.P., Men, H.L.: Scale-sets image analysis. *IJCV* **68**(3), 289–317 (2006)
11. Guimarães, S.J.F., Cousty, J., Kenmochi, Y., Najman, L.: A hierarchical image segmentation algorithm based on an observation scale. In: Gimel'farb, G., Hancock, E., Imiya, A., Kuijper, A., Kudo, M., Omachi, S., Windeatt, T., Yamada, K. (eds.) SSPR & SPR 2012. LNCS, vol. 7626, pp. 116–125. Springer, Heidelberg (2012)
12. Guimarães, S.J.F., Patrocínio Jr., Z.K.G.: A graph-based hierarchical image segmentation method based on a statistical merging predicate. In: Petrosino, A. (ed.) ICIAP 2013, Part I. LNCS, vol. 8156, pp. 11–20. Springer, Heidelberg (2013)
13. Martin, D.R., Fowlkes, C.C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. *PAMI* **26**(5), 530–549 (2004)
14. Morris, O., Lee, M.J., Constantinides, A.: Graph theory for image analysis: an approach based on the shortest spanning tree. *IEEE Proceedings F (Communications, Radar and Signal Processing)* **133**(2), 146–152 (1986)
15. Najman, L.: On the equivalence between hierarchical segmentations and ultrametric watersheds. *JMIV* **40**, 231–247 (2011)
16. Nock, R., Nielsen, F.: Statistical region merging. *PAMI* **26**(11), 1452–1458 (2004)
17. Peng, B., Zhang, D., Zhang, D.: Automatic image segmentation by dynamic region merging. *IEEE Trans. on Image Processing* **20**(12), 3592–3605 (2011)
18. Rother, C., Kolmogorov, V., Blake, A.: “grabcut”: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **23**(3), 309–314 (2004)
19. Zahn, C.T.: Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Trans. Comput.* **20**, 68–86 (1971)