A Graph-Based Hierarchical Image Segmentation Method Based on a Statistical Merging Predicate[★]

Silvio Jamil F. Guimarães and Zenilton K.G. Patrocínio Jr.

Audio-Visual Information Proc. Lab. (VIPLAB)
Computer Science Department – ICEI – PUC Minas
{sjamil,zenilton}@pucminas.br

Abstract. Hierarchical image segmentation provides a set of image segmentations at different detail levels in which coaser details levels can be produced by simple merges of regions from segmentations at finer detail levels. Most image segmentation algorithms, such as region merging algorithms, rely on a criterion for merging that does not lead to a hierarchy. In addition, for image segmentation, the tuning of the parameters can be difficult. In this work, we propose a hierarchical graph-based image segmentation relying on a statistical region merging. 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 two image databases show efficiency and ease of use of our method.

Keywords: hierarchical segmentation, vertex-edge-weighted graph, statistical region merging predicate.

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.

(a) Original   (b) $Q = 16$ (09)   (c) $Q = 34$ (12)   (d) $Q = 35$ (10)
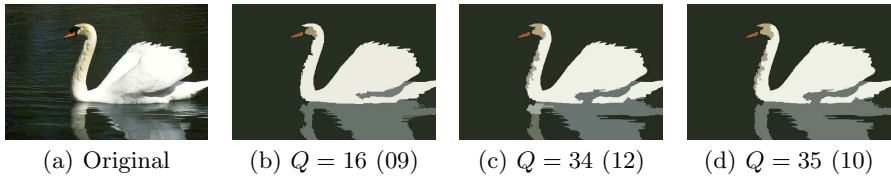
Fig. 1. A real example illustrating the violation of the causality principle by [12]: the number of regions (in parentheses) is not monotonic, when the so-called "segmentation scale" increases

Any hierarchy can be represented with a minimum spanning tree. The first appearance of this tree in pattern recognition dates back to the seminal work of Zahn [13]. Lately, its use for image segmentation was introduced by Morris *et al.* [10] in 1986 and popularized in 2004 by Felzenszwalb and Huttenlocher [5], Noch and Nielsen [12] proposed a statistical method in which the merging order is similar to the creation of a MST. However the region-merging method [5, 12] does not provide a hierarchy. In [4, 11], it was studied some optimality properties of hierarchical segmentations. Considering that, for a given image, one can tune the parameters of the well-known method [5] for obtaining a reasonable segmentation of this image. In [8], we presented a framework to transform the non-hierarchical method proposed by [5] into its hierarchical version, the scale is computed by a maximum between two independent scales. Unfortunately, this methodology can not be directly applied to other methods. For example, the scales computed by the method proposed in [12], here-after called SRG, are dependent on a relation between two regions. Here, we provide a hierarchical version of this method that removes the need for parameter tuning.

Even if the image segmentation results obtained by the method proposed in [12] are interesting, the user faces two major issues:

– first, the number of regions is not monotonic when the parameter k increases. This should not be possible if k was a true scale: indeed, it violates the *causality principle* of multi-scale analysis, that states in our case [6] that a contour present at a scale k_1 should be present at any scale $k_2 < k_1$. Such unexpected behavior of missing causality principle is demonstrated on Fig. 1.
– Second, even when the number of regions decreases, contours are not stable: they can move when the parameter k varies, violating a *location principle*. Such a situation is illustrated on Fig. 2.

Given these two issues, the tuning of the parameters of [5, 12] is a difficult task.

Following [6], we believe that, in order for k to be a true scale-parameter, we have to satisfy both the causality principle and the location principle, which leads to work with a hierarchy of segmentations. Reference [9] is the first to propose an algorithm producing a hierarchy of segmentations based on [5]. However, this method is an iterative version of [5] that uses a threshold function, and requires a tuning of the threshold parameter, therefore this methodology can easily be extended to other methods. Reference [8] is the first to propose a method which

(a) Original   (b) $Q = 16$ (16)   (c) $Q = 32$ (17)   (d) $Q = 64$ (27)
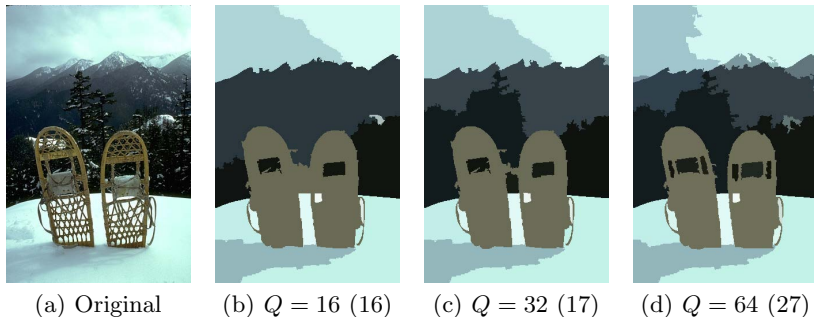
Fig. 2. A real example illustrating the violation of the location principle by [12]: the contour location is not stable for two different "segmentation scales"

computing all hierarchy using the same dissimilarity measure of [5]. However, the extension of this methodology for other methods is not a trivial task.

The main result of this paper is an efficient hierarchical image segmentation algorithm based on the dissimilarity measure of [12]. Our algorithm has a computational cost similar to [12], but provides all statistical scales instead of only one segmentation level. As it is a hierarchy, the result of our algorithm satisfies both the locality principle and the causality principle. Namely, and in contrast with [12], the number of regions is decreasing when the scale parameter increases, and the contours do not move from one scale to another. Therefore, [12] produces under-segmented regions and our approach produces over-segmented regions since small regions may have high hierarchical values.

This work is organized as follows. In Section 2, we present the statistical predicate for merging two regions. In Section 3, we present our hierarchical method for color image segmentation. Some experimental results performed on a image database are given in Section 4. Finally, in Section 5, some conclusions are drawn and further works are discussed.

2   A Statistical Region Merging Predicate

Noch and Nielsen [12] proposed an efficient method for image segmentation based on a statistical predicate. Even if the statistical theory is not easy, the method is simple coping with noise corruption and handle occlusion. To follow, we describe this method.

Let I be an image, $|I|$ be its number of pixels, Q and g be the new and old number of expected different color values. Let $\mathcal{R}_l$ be the set of regions with l pixels and $b(R) = g\sqrt{(1/(Q|R|))ln(|\mathcal{R}_{|\mathcal{R}|}|/\delta)}$ in which $\delta = \frac{1}{(6|I|^2)}$. Thus,

$$b(R, R') \leq \sqrt{b^2(R) + b^2(R')} < b(R) + b(R') \tag{1}$$

According to [12], a region is an unordered bag of pixels, thus we may fix $|\mathcal{R}_l| \leq (l + 1)^{\min\{l,g\}}$. For merging two adjacent regions R and R' with color

average $|R|$ and $|R'|$, respectively, it is necessary to verify the following the statistical region merging predicate

$$\mathcal{P}(R, R') = \begin{cases} \text{true} & \textit{if } |\overline{R} - \overline{R'}| \leq \sqrt{b^2(R) + b^2(R')} \\ \text{false} & \textit{otherwise} \end{cases} \quad (2)$$

The algorithm proposed by [12] for image segmentation considers the statistical region merging predicate defined by Eq. 2 for merging two adjacent regions. To follow, their method is summarized. In 4-connexity, there are $N < 2|I|$ couples of adjacent pixels. Let S_f be the set of these couples. Let $f(p, p')$ be a real-valued function, with p and p' pixels of f. First, the couples of S_f are sorted in increasing order of $f(.,.)$, and then traverse this order only once. For any current couple of pixels p and p' that does not in the same region, the test $\mathcal{P}(R, R')$, and merge R and R' iff it returns true.

Considering that the parameter Q controls the coarseness (or "scale") of a segmentation, it is important to understand the influence of this parameter in the process. Thus, we re-write the Eq. 2 in order to explicit the parameter Q:

$$(\overline{R} - \overline{R'})^2 \leq b^2(R) + b^2(R') \quad (3)$$

$$(\overline{R} - \overline{R'})^2 \leq g^2 \frac{1}{(Q|R|)} ln \frac{|\mathcal{R}_{|\mathcal{R}|}|}{\delta} + g^2 \frac{1}{(Q|R'|)} ln \frac{|\mathcal{R}_{|\mathcal{R'}|}|}{\delta} \quad (4)$$

$$(\overline{R} - \overline{R'})^2 \leq g^2 \frac{|R'|}{(Q|R||R'|)} ln \frac{|\mathcal{R}_{|\mathcal{R}|}|}{\delta} + g^2 \frac{|R|}{(Q|R'||R|)} ln \frac{|\mathcal{R}_{|\mathcal{R'}|}|}{\delta} \quad (5)$$

$$Q \leq \frac{g^2}{(\overline{R} - \overline{R'})^2} \left(\frac{|R'|}{(|R||R'|)} ln \frac{|\mathcal{R}_{|\mathcal{R}|}|}{\delta} + \frac{|R|}{(|R'||R|)} ln \frac{|\mathcal{R}_{|\mathcal{R'}|}|}{\delta} \right) \quad (6)$$

Thanks to upperbound of $|\mathcal{R}_l|$, the Eq. 6 can be re-written by

$$Q \leq \frac{g^2}{(\overline{R} - \overline{R'})^2 |R||R'|} (|R'| \, m \, (l + d) + |R| \, m' \, (l' + d)) \quad (7)$$

in which $m = \min\{g, |R|\}$, $m' = \min\{g, |R'|\}$, $l = ln(1 + |R|)$, $l' = ln(1 + |R'|)$ and $d = 2 \times \, ln(6|I|)$.

We can see in Fig. 2 and Fig. 1 the segmentation results for fixed values of parameter Q (which controls the so-called *segmentation scale*), however the location and causality principle are missing, and this shows the absence of hierarchical properties of this method. In fact, we will look for to adapt the values of Q according to the regions analyzed for guaranteeing that two regions are correctly merged. First, since Q could easily be limited by the number of pixels in the image, i.e., $Q \leq |I|$, we consider in our analysis $Q' = |I| - Q$ because minimizing Q' will correspond to maximizing Q (we search for the smallest scale value for which two regions can be merged).

Then, the *scale* $Q(R, R')$ is defined as:

$$Q(R, R') = |I| - \frac{g^2}{(\overline{R} - \overline{R'})^2 |R||R'|} (|R'| \, m \, (l + d) + |R| \, m' \, (l' + d)). \quad (8)$$

Thanks to this notion of a scale, Eq. (7) can be written as:

$$Q' \geq Q(R, R'). \tag{9}$$

3   A Hierarchical Graph Based Image Segmentation

In this section, we describe our method to compute a hierarchy of partitions based on scales, so-called here *statistical scale*, as defined by Eq. 7. This method is based on our previous work [7, 8], however here it is necessary to compute all candidate values to represent a scale once, differently of [8], the merging depends on both, the information of the regions and the distance between these regions. Let us first recall some important notions for handling hierarchies [4, 10, 11].

To every tree T spanning the set V of the image pixels, to every map $w : E \to \mathbb{N}$ that weights the edges of T and to every threshold $\lambda \in \mathbb{N}$, one may associate the partition $\mathcal{P}_\lambda^w$ of V induced by the connected components of the graph made from V and the edges of weight below λ. It is well known [4, 10] that for any two values λ_1 and λ_2 such that $\lambda_1 \geq \lambda_2$, the partitions $\mathcal{P}_{\lambda_1}^w$ and $\mathcal{P}_{\lambda_2}^w$ are *nested* and $\mathcal{P}_{\lambda_1}^w$ is *coarser* than $\mathcal{P}_{\lambda_2}^w$. Hence, the set $\mathcal{H}^w = \{\mathcal{P}_\lambda^w \mid \lambda \in \mathbb{N}\}$ is a *hierarchy of partitions induced by the weight map* w.

Our algorithm does not explicitly produce a hierarchy of partitions, but instead produces a weight map L (scales of statistical values) from which the desired hierarchy $\mathcal{H}^L$ can be inferred on a given T. It starts from a minimum spanning tree T of the vertex-edge-weighted graph built from the image. In order to compute the scale $L(e)$ associated with each edge of T, 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 map $L(e)$ is initialized to $|I|$; then for each edge e linking two vertices p and p' the following steps are performed:

(i) Find the region R of $\mathcal{P}_{w(e)}^w$ that contains p.
(ii) Find the region R' of $\mathcal{P}_{w(e)}^w$ that contains p'.
(iii) Compute the hierarchical statistical scale $L(e) = Q(R, R')$.

At step (iii), the *hierarchical statistical scale* $L(e)$ is computed based on a minimization of the function which relates some sub-region of R to some sub-region of R'. More precisely, using an internal parameter v, this scale is computed as follows:

(1) Initialize the value of v to $|I|$.
(2) Decrement the value of v by 1.
(3) Find the region R_* of $\mathcal{P}_v^L$ that contains p.
(4) Find the region R'_* of $\mathcal{P}_v^L$ that contains p'.
(5) Repeat steps 2, 3 and 4 while $Q(R_*, R'_*) < v$
(6) Set $Q(R, R') = v$.

As the image is represented by a vertex-edge-weighted graph and the dissimilarity measures is based on color information of the regions, the computation of

color average for each region can be done by computation of average cost of the region vertices.

To efficiently implement our method, we use some data structures similar to the ones proposed in [4, 8]; in particular, the management of the collection of partitions is due to Tarjan's union find and Fredman and Tarjan's Fibonnacci heaps. Furthermore, we made some algorithmic optimizations to speed up the computations of the statistical scales.

4   Experiments

To provide a basis of comparison for our hierarchical method with respect to its non-hierarchical version, we consider the dataset used in [1, 2] which contains clearly one or two objects. This dataset is divided into two groups containing 100 images each one: single and two objects, this dataset is so-called **Object data set**. According to [1, 2] the database was designed to contain a variety of images with objects that differ from their surroundings by either intensity, texture, or other low level cues. To avoid potential ambiguities it was selected images that clearly depict one object or two objects in the foreground. Some experiments are also given using Berkeley Segmentation DataSet [3], so-called **BSDS500**. In this database, a semantic segmentation is done, and sometimes, some images are under-segmented. With respect to the quality measures, we assess our method using measures presented in [3] (Ground-Truth Segmentation, Probabilistic Rand Index and Variation Information) and measures used in [1, 2] (F1-measure). The aim here is to show how the inclusion of hierarchical property will influence on the results obtained by the method without this property.

In Table 1 and 2, we present some results, represented by F-measures, obtained when we apply the method to Object Data set. As we can observe, the F-measure calculated for the results obtained by hSRG are much better than the

Table 1. Performances of our methods, called hSRG and hFH, when compared to their non-hierarchical method, SRG and FH, respectively. The performances showed in [2] are also presented. The performances are measures by F-measure for the Single-Object Data Set: (i) single segment coverage; and (ii) fragmented Coverage. See [2] for more details on the evaluation method.

Algorithm	Single	Multi	
	F-measure	F-measure	Fragmentation
Proba	0.86	0.87	2.66
SWA V1	0.83	0.89	3.92
SWA V2	0.76	0.86	3.71
Gpb	0.54	0.88	8.20
Mean Shift	0.57	0.88	12.08
NCut	0.72	0.84	3.12
SRG	0.58	0.56	8.53
hSRG	0.64	0.88	7.83
FH	0.73	0.85	5.96
hFH	0.63	0.85	5.12

one obtained by SRG. For hFH and FH, the measures are quite similar. Thus, these results illustrate that the inclusion of hierarchical property has no prejudice with respect to the computed measures.

Table 2. Performances of our methods, called hSRG and hFH, when compared to their non-hierarchical method, SRG and FH, respectively. The performances showed in [2] are also presented. The performances are measures by F-measure for fragmented coverage for the Two-Object Data Set: (i) fragmented coverage for smaller object; and (ii) fragmented coverage for larger object. See [2] for more details on the evaluation method.

(a) Fragmented coverage

Algorithm	2 Objects		Larger object		Smaller object	
	Aver. F-measure	Aver. Frag.	Aver. F-measure	Aver. Frag.	Aver. F-measure	Aver. Frag.
Proba	0.85	1.67	0.87	2.00	0.84	1.33
SWA V1	0.88	3.13	0.91	3.88	0.84	2.37
SWA V2	0.85	2.27	0.88	2.76	0.82	1.77
Gpb	0.84	2.95	0.87	3.60	0.81	2.30
Mean Shift	0.78	3.65	0.85	4.49	0.71	2.81
NCut	0.84	2.64	0.88	3.34	0.80	1.93
SRG	0.60	1.90	0.60	1.68	0.60	2.12
hSRG	0.87	4.96	0.88	4.77	0.87	5.15
FH	0.88	5.86	0.89	6.91	0.87	3.81
hFH	0.86	4.20	0.89	5.60	0.84	2.81

(b) Single coverage

Algorithm	2 Objects	Larger object	Smaller object
Proba	0.68	0.70	0.65
SWA V1	0.66	0.74	0.57
SWA V2	0.61	0.71	0.50
Gpb	0.72	0.70	0.75
Mean Shift	0.61	0.65	0.58
NCut	0.58	0.66	0.49
SRG	0.58	0.62	0.55
hSRG	0.75	0.74	0.76
FH	0.79	0.78	0.80
hFH	0.75	0.76	0.75

In Table 3, we present the quality measures when we apply four segmentation methods to the object dataset: two non-hierarchical methods (SRG and FH) and their hierarchical versions, called hSRG and hFH). In this experiment, we compute all measures for the datasets containing one object and two objects, we also compute the quality measures considering that we search only the smaller or larger objects into the two-object dataset. As we can see, these results are quite similar.

Table 3. Performances of hierarchical methods, called hSRG and hFH [8], and the non-hierarchical methods SRG [2] and FH [5] using three different measures applied to **Object Data Set**: (a) Ground-truth Covering (GT Covering), (b) Probabilistic Rand Index and (c) Variation Information. The presented scores are optimal considering a constant scale parameter for the whole dataset (ODS) and a scale parameter varying for each image (OIS), and a scale parameter varying for each region of the ground-truth (Best). See [3] for more details on the evaluation method.

(a) Ground-truth Covering (GT Covering)

	Smaller			Large			2 obj			1 obj		
	ODS	OIS	Best	ODS	OIS	Best	ODS	OIS	Best	ODS	OIS	Best
SRG	0.82	0.83	0.85	0.80	0.83	0.85	0.78	0.85	0.87	0.67	0.73	0.75
hSRG	0.81	0.84	0.88	0.80	0.84	0.88	0.73	0.80	0.84	0.65	0.73	0.76
FH	0.81	0.84	0.88	0.80	0.83	0.87	0.73	0.83	0.87	0.66	0.73	0.77
hFH	0.82	0.84	0.89	0.82	0.85	0.88	0.75	0.83	0.86	0.63	0.67	0.70
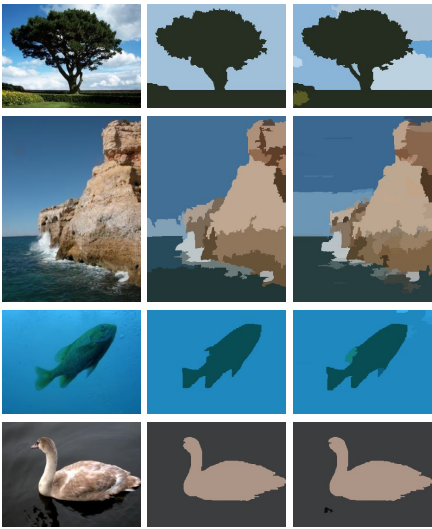
(b) Probabilistic Rand Index

	Smaller		Large		2 obj		1 obj	
	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS
SRG	0.80	0.81	0.79	0.83	0.80	0.86	0.70	0.77
hSRG	0.80	0.83	0.80	0.83	0.74	0.83	0.68	0.78
FH	0.80	0.83	0.78	0.82	0.73	0.84	0.68	0.77
hFH	0.81	0.84	0.81	0.84	0.76	0.86	0.65	0.75
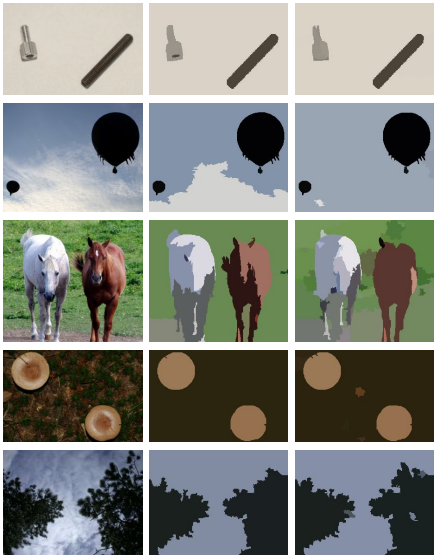
(c) Variation Information

|	Smaller		Large		2 obj		1 obj	
	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS
SRG	0.60	0.57	0.63	0.60	0.69	0.57	0.98	0.91
hSRG	0.57	0.54	0.58	0.55	0.84	0.72	0.85	0.80
FH	0.55	0.52	0.59	0.55	0.77	0.60	0.96	0.87
hFH	0.56	0.56	0.54	0.52	0.77	0.64	0.92	0.89

(a)

(b)

Fig. 3. Examples of application of SRG and hSRG methods to Object Dataset containing one object and two objects. In left column of (a) and (b), we present the original images, in middle column, we illustrate the results obtained from SRG method. In right column, we illustrate the results obtained from hSRG.
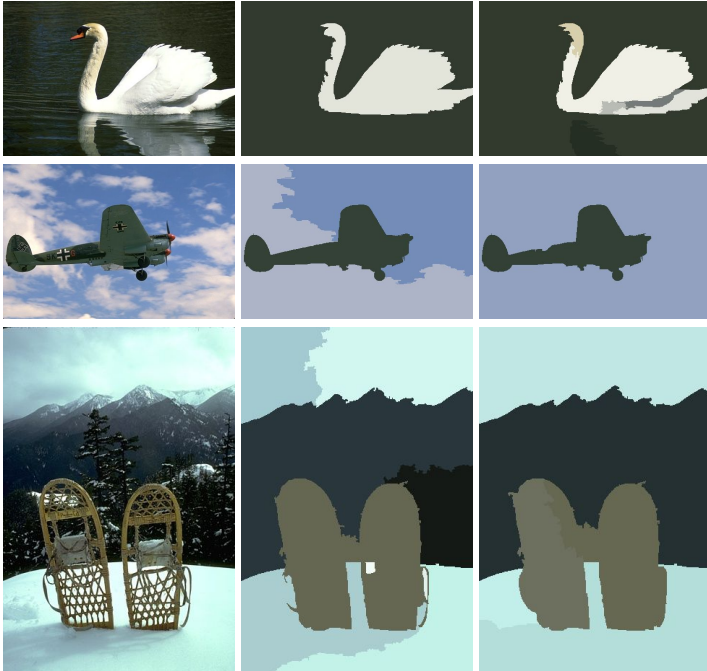
Fig. 4. Examples of application of SRG and hSRG methods to BSDS. In left column, we present the original images, in middle column, we illustrate the results obtained from SRG method. In right column, we illustrate the results obtained from hSRG.

In Fig. 3 and 4 we present some results of applications of SRG method and our hierarchical version. The two formers were extracted from Object Data Set, and the last one, is obtained from BSDS500. As we can see, the results are quite similar, however there are some differences, mainly, in second and third rows of Fig. 4, in second and third rows in Fig 3(b) and in first row of Fig. 3(a).

5   Conclusion and Further Works

In this work, we propose a method for transforming a non-hierarchical method into a hierarchical method preserving the merging criterium, *i.e.*, all regions are merged according to the same statistical criterium. Differently of the method that iteratively computes the hierarchies, our method produces a weight map L (scales of statistical 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 location problems which are missing in SRG method, and does not prejudice the quality of results, in fact, our method present higher f-measure values than SRG method. However, due to inclusion of hierarchical property, some small regions are present in high

scale. To solve this problem, it is necessary to filter out this regions using an area opening. For future works, we will study the robustness to noise and how to automatically choice 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: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (June 2007)
2. Alpert, S., Galun, M., Brandt, A., Basri, R.: Image segmentation by probabilistic bottom-up aggregation and cue integration. IEEE Trans. Pattern Anal. Mach. Intell. 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. 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)
5. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. IJCV 59, 167–181 (2004),
   `http://portal.acm.org/citation.cfm?id=981793.981796`
6. Guigues, L., Cocquerez, J.P., Men, H.L.: Scale-sets image analysis. IJCV 68(3), 289–317 (2006), `http://dx.doi.org/10.1007/s11263-005-6299-0`
7. Guimarães, S.J.F., Cousty, J., Kenmochi, Y., Najman, L.: An efficient hierarchical graph based image segmentation. CoRR abs/1206.2807 (2012)
8. 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)
9. Haxhimusa, Y., Kropatsch, W.: Segmentation graph hierarchies. In: Fred, A., Caelli, T.M., Duin, R.P.W., Campilho, A.C., de Ridder, D. (eds.) SSPR&SPR 2004. LNCS, vol. 3138, pp. 343–351. Springer, Heidelberg (2004)
10. Morris, O., Lee, M.J., Constantinides, A.: Graph theory for image analysis: an approach based on the shortest spanning tree. Communications, Radar and Signal Processing IEE Proceedings F 133(2), 146–152 (1986)
11. Najman, L.: On the equivalence between hierarchical segmentations and ultrametric watersheds. JMIV 40, 231–247 (2011)
12. Nock, R., Nielsen, F.: Statistical region merging. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(11), 1452–1458 (2004)
13. Zahn, C.T.: Graph-theoretical methods for detecting and describing gestalt clusters. IEEE Trans. Comput. 20, 68–86 (1971),
    `http://dx.doi.org/10.1109/T-C.1971.223083`