HAL
open science

Hierarchy-Based Salient Regions: A Region Detector Based on Hierarchies of Partitions

Karla Otiniano-Rodríguez, Arnaldo de A. Araújo, Guillermo Cámara-Chávez, Jean Cousty, Silvio Jamil F. Guimarães, Benjamin Perret

To cite this version:

Karla Otiniano-Rodríguez, Arnaldo de A. Araújo, Guillermo Cámara-Chávez, Jean Cousty, Silvio Jamil F. Guimarães, et al.. Hierarchy-Based Salient Regions: A Region Detector Based on Hierarchies of Partitions. CIARP 2018: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, , pp.444-452, 2019, 10.1007/978-3-030-13469-3_52 . hal-02064718

HAL Id: hal-02064718

https://hal.science/hal-02064718v1

Submitted on 12 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hierarchy-Based Salient Regions: a region detector based on hierarchies of partitions*

Karla Otiniano-Rodríguez[1,2], Arnaldo de A. Araújo[1,2], Guillermo Cámara-Chávez[3], Jean Cousty[1], Silvio Jamil F. Guimarães[1,4], and Benjamin Perret[1]

[1] University Paris-Est, LIGM, A3SI, ESIEE, France
[2] Universidade Federal de Minas Gerais, Computer Science Department, Brazil
[3] Universidade Federal de Ouro Preto, Computer Science Department, Brazil
[4] Pontifical Catholic University of Minas Gerais, Computer Science Department, Belo Horizonte-MG, Brazil

Abstract. This article introduces a novel region detector based on hierarchies of partitions, so-called Hierarchy-Based Salient Regions (HBSR). This approach enables to combine the clues given by a high quality contour detector with a custom salient region detection procedure. The evaluation of the proposed method HBSR with a standard feature detection assessment framework shows that HBSR outperforms the state-of-the-art methods, in average. These promising results may lead to improvements in many computer vision tasks.

Keywords: Region detector · Mathematical Morphology · Hierarchy of Partitions · Computer Vision

1 Introduction

The extraction of local image features is a conventional approach for providing compact image descriptors that can be used to solve many computer vision tasks, like image stitching, tracking, reconstruction, image retrieval. Some examples of local features are edges, corners, ridges and blobs. The desirable qualities of image features (*e.g.*, repeatability, distinctiveness, accuracy) [13] are tightly linked to the invariance properties of the detector (*e.g.*, invariance to viewpoint, to luminosity, and to compression). Some of the best-known feature detectors are SIFT [5], SURF [1], ORB [12], MSER [6], Harris-Affine and Hessian-Affine [9]. In this article, we present a local region detector based on hierarchies of partitions.

Existing feature detection methods based on hierarchies, like MSER [6], TBMR [14], or TOS-MSER [2], rely on component trees (min-tree, max-tree, and level-line tree) and thus on the study of the lightness of the image, seen as a topographical relief. Here, we propose to replace the use of component trees by hierarchies of partitions whose construction rely on the gradient of the image.

* This work received funding from FONDECYT-CONCYTEC (contract number 004-2016-FONDECYT), CAPES (PVE 125000/2014-00), FAPEMIG (PPM 00006-16), and CNPq (Universal 421521/2016-3 and PQ 307062/2016-3).

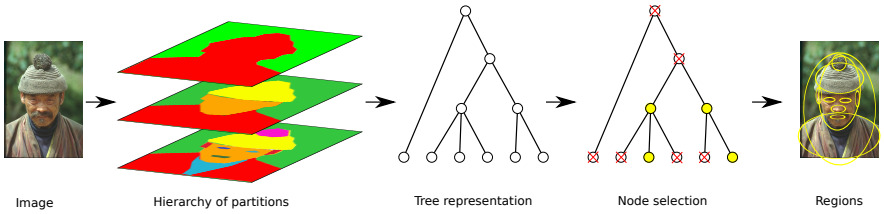

Fig. 1. Main steps of the proposed region detector HBSR.

Actually, this approach allows us to take advantage of machine learning based contour detectors to obtain a high-quality multiscale representation of the image from which we select salient nodes. The evaluation of the proposed method, called *Hierarchy-based Salient Regions* (HBSR), with a standard feature detection assessment framework shows that the proposed method outperforms the current state-of-the-art on average.

This article is organized as follows. Section 2 presents the proposed method and the fundamentals of hierarchy of partitions. Section 3 describes the evaluation framework used in Section 4 for the comparison with the state-of-the-art methods. Finally, conclusions and future works are drawn in Section 5.

2 The novel region detector

Ideally, in a hierarchy of partitions of an image, the scene is iteratively refined in its objects, parts of the objects, parts of the parts, and so on. Thus, each region (also called node) of the hierarchy should represent a *salient* element of the scene. However, in practice, hierarchical representations are not perfect and generally contain artifacts (regions that do not correspond to any meaningful element of the scene) and redundancy (several nodes representing the same region with slight variations). The proposed method aims at selecting nodes from a hierarchy of partitions of an image by determining the *salient nodes* of the hierarchy and then filtering redundancy among them (see Figure 1). Finally, each selected node of the hierarchy is represented by its best fitting ellipse.

2.1 Preliminary definitions

In the sequel of this article, the *graph* $\mathcal{G}$ is defined as a pair (V, E) where V is a finite set and E is composed of pairs of distinct elements in V, *i.e.*, E is a subset of $\{\{x, y\} \subseteq V \mid x \neq y\}$. Each element of V is called a *vertex or a pixel* (of $\mathcal{G}$), and each element of E is called an *edge* (of $\mathcal{G}$). The graph $\mathcal{G}$ provides a structure to the image spatial domain, *i.e.*, V is the regular 2D grid of pixels, and E is the 4- or 8-adjacency relation. We denote by W a function from E to $\mathbb{R}$ that weights the edges of $\mathcal{G}$. Therefore, the pair $(\mathcal{G}, W)$ is an *edge-weighted graph*, and, for any $u \in E$, the value $W(u)$ is the *weight of* u.

A *hierarchy* (or *dendrogram*) $\mathcal{H}$ of $\mathcal{G}$ is a family of subsets of V such that any two elements A and B of $\mathcal{T}$ are either nested or disjoint: *i.e.*, $A \cap B \in \{\emptyset, A, B\}$.

Any element of $\mathcal{H}$ is called a *node* or *region* of $\mathcal{H}$. The minimal elements of $\mathcal{H}$ are called the *leaves*. The parent of a node $N \neq V$ of $\mathcal{H}$, denoted by $Parent(N)$, is the smallest node N' of $\mathcal{H}$ that is strictly larger than N. Conversely, we say that a node N is a child of its parent $Parent(N)$. When the leaves, *i.e.*, the nodes without any child, of the hierarchy $\mathcal{H}$ forms a partition of V, then the hierarchy can be represented as a sequence of nested partitions (see Figure 1).

2.2 Selection of salient regions

We aim at selecting the salient regions from a hierarchy $\mathcal{H}$ obtained from the weighted graph $(\mathcal{G}, W)$. The result of this selection process is a new hierarchy $\mathcal{H}'$ whose nodes are the selected regions of $\mathcal{H}$. Salient regions are identified based on three local features: size, contrast, and geometrical complexity. In the following of this section, R denotes a region of the hierarchy $\mathcal{H}$.

Size criterion The *area* of the region R, denoted by $A(R)$, is defined as the number of vertices in R (*i.e.*, $A(R) = |R|$). We assume that a salient region is neither too small nor too large, leading to the following selection criterion: $A_{min} \leq A(R) \leq A_{max}$, with A_{min} and A_{max} two real parameters representing respectively the minimum and maximum area of a salient region.

Contrast criterion We consider that the edge-weights of the graph represent gradient values between pixels. The contrast being a relative measure of difference between the region and its surroundings, we use the gradient inside the parent of the given region to estimate it. We define the *depth* of the region R, denoted by $D(R)$, as the maximal weight of the edges linking two vertices of the parent region of R (*i.e.*, $D(R) = \max\{W(e), e \in E \mid e \subseteq Parent(R)\}$). We assume that a salient region should have a significant contrast leading to the following criterion: $D_{min} \leq D(R)$, with D_{min} a real parameter representing the minimum depth of a salient region.

Shape complexity criterion The ellipse is a common shape used to represent a region in an image [15], and a way to measure the geometric complexity of a region is to quantify the difference between the real shape and its best fitting ellipse. We define the *shape complexity* of R, denoted $C(R)$, as the ratio of the area of the best fitting ellipse of R (estimated with second ordered moments), denoted by $A(E_R)$, with the area of R (*i.e.*, $C(R) = A(E_R)/A(R)$). We assume that a salient region should have a low shape complexity leading to the following criterion: $C(R) \leq C_{max}$, with C_{max} a real parameter representing the maximum shape complexity of a salient region.

Thus, we use these criteria for identifying candidate regions on a given hierarchy of partitions $\mathcal{H}$. The result is a new hierarchy $\mathcal{H}_1$ composed of the regions of $\mathcal{H}$ identified as salient:

$$\mathcal{H}_1 = \{R \in \mathcal{H} \mid A_{min} \leq A(R) \leq A_{max},\ D_{min} \leq D(R), \text{ and } C(R) \leq C_{max}\}.$$

2.3 Filtering of redundant regions

The new hierarchy $\mathcal{H}_1$ composed by the salient regions of $\mathcal{H}$ may still contain redundant regions, *i.e.*, very similar nodes. The aim of the filtering procedure

presented in this section is to select a representative node from similar ones. Thus, we propose a two-step procedure to perform this selection:

- Similarly to [14], we identify topological changes in the hierarchy as regions having at least two children. Indeed, when a region of the hierarchy has a single child, it cannot be viewed as the decomposition of an object into its parts. Therefore, the single child of this region is discarded. Formally, this process leads to a new hierarchy $\mathcal{H}_2$ defined by:

$$\mathcal{H}_2 = \{R \in \mathcal{H}_1 \mid Ch(Parent(R)) \geq \ 2\},$$

  where $Ch(Parent(R))$ is the number of children of the parent region of R.
- Then, we discard a node when its shape is similar to the one of its parent. The dissimilarity between the shapes of two regions is evaluated by computing the relative difference of the area of their best fitting ellipses. This leads to a final hierarchy $\mathcal{H}_3$:

$$\mathcal{H}_3 = \left\{ R \in \mathcal{H}_2 \;\middle|\; \frac{|A(E_R) - A(E_{Parent(R)})|}{A(E_{Parent(R)})} \geq DS_{min} \right\},$$

  where DS_{min} is a real parameter representing the minimum dissimilarity between a region and its parent.

The final set of detected regions is composed of the best fitting ellipses of the regions of $\mathcal{H}_3$. Regarding the computational cost, the detection of salient regions and the filtering of the redundant regions can be computed in linear time with respect to the number of vertices in the graph $\mathcal{G}$.

3 Evaluation framework

We rely on the framework of Mikolajczyk et al. [8] to provide an objective assessment of the proposed method. The framework is associated with a dataset of eight image sequences, with six images each. The dataset includes five types of transformations: viewpoint changes (a) & (b); scale changes (c) & (d); image blur (e) & (f); JPEG compression (g); and illumination (h) (see Figure 2).

For each image sequence of the dataset, the framework compares the regions provided by the detectors on the first image of the sequence with the ones obtained on the other images of the sequence. Two measures are used as follows:

1. the *repeatability score* which evaluates the theoretical performance of the detector by calculating the ratio of the number of correspondences between regions of the two images and the number of proposed regions. Given two regions, we say that there is a correspondence if the overlap error between their best fitting ellipses is small; and
2. the *matching score* which evaluates the practical performance of the detector by calculating the ratio of the number of correct matches in the feature space and the number of proposed regions. A match between two regions is considered correct if they are nearest neighbours in the feature space, and if they have the smallest overlap error.

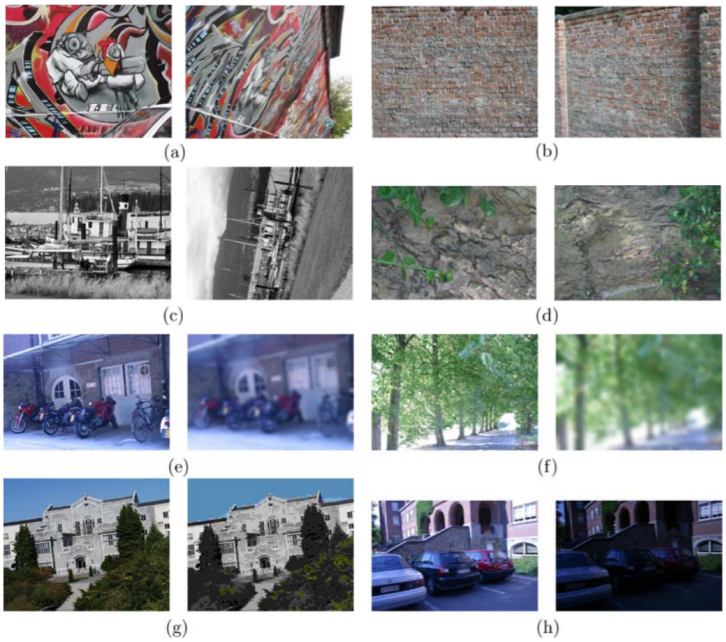

Fig. 2. Some examples for each sequence of the dataset. (a) Graffiti, (b) Wall, (c) Boat, (d) Bark, (e) Bikes, (f) Trees, (g) UBC and (h) Leuven.

4 Experimental analysis

In this section, we discuss the experimental results showing some illustrations of our region detector and the quantitative comparison between the proposed method HBSR and the state-of-the-art methods.

4.1 Experimental setup

In the following experiments, an image is represented as a 4-adjacency graph from which a *Quasi-Flat Zones (QFZ) hierarchy* [7] is computed. QFZ hierarchies are naturally invariant to photometric changes and geometric changes (up to quantization effects). A quasi-flat zone of the weighted graph $(\mathcal{G}, W)$ at level $\lambda \in \mathbb{R}$ is a maximal set of vertices such that, between any two of its vertices, there exists a path along which the maximal weight is λ. The set of quasi-flat zones of the weighted graph at all levels λ forms the *quasi-flat zones hierarchy* of the weighted graph. According to [11], we chose to use the Structured Edge Detector (SED) [4] in order to weight the edges of the graph: indeed this detector offers good performances in combination with quasi-flat zones hierarchies on natural images while being fast to compute. To further improve the invariance of the salient region detection process (in particular, the definition of the depth of a region), we propose to perform a histogram normalization of the gradient

produced by SED. Note that, the QFZ hierarchy can be efficiently computed in (quasi) linear time from the graph weighted by SED [10, 3].

The proposed region detector has five parameters, which were optimized to maximize the average of the repeatability and matching scores on the evaluation dataset: the minimum area ($A_{min} = 0.08$), the maximum area ($A_{max} = 0.25$), the minimal depth ($D_{min} = 22$), the maximal shape complexity ($C_{max} = 1.1$), and the minimum dissimilarity ($DS_{min} = 20\%$). Note that the area parameter are expressed as a percentage of the total image size.

4.2 Quantitative assessment

In this section, we assess the proposed method HBSR within Mikolajczyk et al. framework [8]. We provide quantitative results and a discussion about the invariance of our method, against geometric and photometric changes, by analyzing the results of each sequence of the dataset separately. The proposed method is compared to four state-of-the-art region detectors: Harris-Affine [9], Hessian-Affine [9], Maximally Stable Extremal Region (MSER) [6], and Tree-Based Morse Regions (TBMR) [14]. The Harris-Affine and the Hessian-Affine are two related methods which detect interest points in scale-space based on the Laplacian operator. The MSER and TBMR detectors both operate on hierarchical representations of the images called min- and max-trees that represent the minima (respectively maxima) of the image and their merging order as the brightness increases (respectively decreases). While MSER looks for long branches of the hierarchy with small area variations, TBMR searches for topological changes (critical points of the lightness function) in the hierarchy. Figure 3 shows the regions provided by our detector on some images of the evaluation

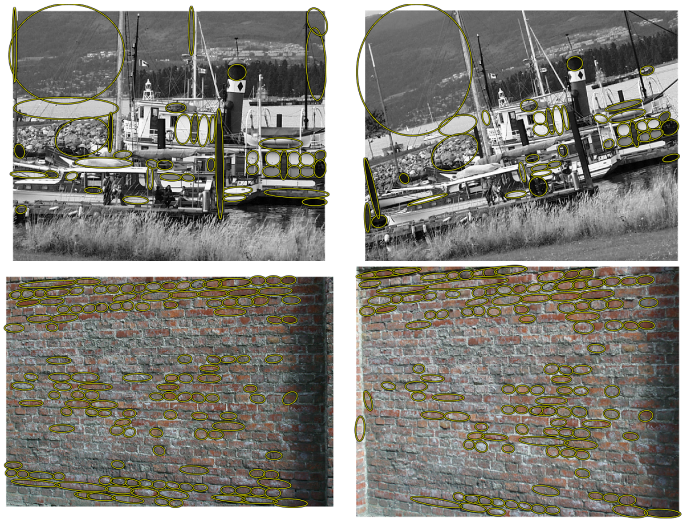

Fig. 3. Detected regions (yellow ellipses) by HBSR, the proposed region detector, on images from the *Boat* and the *Wall* sequences.

Table 1. Repeatability and matching scores. Each value represents the average repeatability of the five results of each detector for each sequence.

	Sequence	Harris	Hessian	MSER	TBMR	HBSR
Repeatability	**Graffiti**	39.69	47.10	**64.48**	52.07	41.13
	Wall	44.63	49.17	55.13	53.63	**64.42**
	Boat	49.28	**59.21**	50.69	42.04	49.73
	Bark	60.94	**75.98**	48.73	63.03	30.00
	Bikes	56.52	73.25	57.12	50.46	**87.35**
	Trees	50.25	55.39	39.61	40.24	**70.00**
	UBC	80.37	**86.12**	50.01	45.69	71.11
	Leuven	56.27	67.50	**77.85**	68.50	76.78
	Average	54.74	**64.22**	55.46	51.96	61.32
Matching	**Graffiti**	16.26	20.45	**48.31**	36.96	28.50
	Wall	19.33	26.63	37.90	34.35	**42.08**
	Boat	24.68	31.04	35.49	27.83	**41.89**
	Bark	33.38	**45.67**	26.16	38.43	30.00
	Bikes	30.62	38.29	42.27	28.09	**82.46**
	Trees	9.02	12.78	17.11	11.55	**70.00**
	UBC	64.10	**70.44**	41.28	27.91	66.67
	Leuven	30.55	34.93	**65.59**	52.83	64.50
	Average	28.49	35.03	39.27	32.24	**53.26**
Average score		41.62	49.63	47.37	42.10	**57.29**

dataset. We can see that the proposed detector produces a reasonable number of regions corresponding to well identified shapes of the scene.

Table 1 shows the results of repeatability and matching scores. The results obtained on each sequence are presented separately in order to analyze the results of each geometrical or photometrical change. We can observe that HBSR is particularly robust to blurring (*Bikes* and *Trees* sequences) where it obtains best repeatability and matching scores. Luminosity changes (*Leuven* sequence) and JPEG compression artifacts (*UBC* sequence) are also very well handled with repeatability and matching scores very close to the ones. The proposed method also manages to deal with moderate viewpoint change on highly textured images (*Wall* sequence) very well (first on both scores). Significant viewpoint changes (*Graffiti* and *Boat* sequences) are however moderately well handled with average scores. Finally, the main weakness of the proposed method appears with large viewpoint changes combined with smooth surfaces (*Bark* sequence) where the SED contour detector fails to detect any meaningful contour, hence leading to the absence of meaningful regions. Furthermore, Table 1 also shows aggregated repeatability and matching scores in terms of average on the eight sequences. We can see that our method obtains the best average score, with an average repeatability very close to the best method and with an average matching score significantly higher than all other methods.

5 Conclusion

We presented HBSR, a local region detector based on hierarchies of partitions, that allows us to take advantage of high-quality contour detectors. We proposed several heuristics to select and filter redundant regions from a hierarchy of partitions to obtain robust, relevant and multi-scale regions of an image. Our experiments show promising results, with better average results than state-of-the-art methods. In future works, we plan to improve the node selection method further, to experiment with other hierarchies of partitions, and to apply the proposed method to various computer vision tasks.

References

1. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (surf). CVIU **110**(3), 346–359 (2008)
2. Bosilj, P., Kijak, E., Lefèvre, S.: Beyond MSER: Maximally Stable Regions using Tree of Shapes. In: BMVC. Swansea, United Kingdom (2015)
3. Cousty, J., Najman, L., Perret, B.: Constructive links between some morphological hierarchies on edge-weighted graphs. In: ISMM. pp. 86–97 (2013)
4. Dollár, P., Zitnick, C.L.: Fast edge detection using structured forests. IEEE TPAMI **37**(8), 1558–1570 (2015)
5. Lowe, D.G.: Object recognition from local scale-invariant features. In: ICCV. pp. 1150–1157 (1999)
6. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. IVC **22**(10), 761–767 (2004)
7. Meyer, F., Maragos, P.: Morphological scale-space representation with levelings. In: Scale-Space Theories in Computer Vision, LNCS, vol. 1682, pp. 187–198 (1999)
8. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. IJCV **65**(1), 43–72 (2005)
9. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. IJCV **60**(1), 63–86 (2004)
10. Najman, L., Cousty, J., Perret, B.: Playing with kruskal: algorithms for morphological trees in edge-weighted graphs. In: ISMM. pp. 135–146. Springer (2013)
11. Perret, B., Cousty, J., Guimarães, S.J., Maia, D.S.: Evaluation of hierarchical watersheds. IEEE TIP **27**(4), 1676–1688 (2018)
12. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.R.: ORB: an efficient alternative to SIFT or SURF. In: ICCV. pp. 2564–2571. IEEE Computer Society (2011)
13. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: A survey. Found. Trends Comput. Graph. Vis. **3**(3), 177–280 (2008)
14. Xu, Y., Monasse, P., Géraud, T., Najman, L.: Tree-Based Morse Regions: A Topological Approach to Local Feature Detection. IEEE TIP **23**(12), 5612–5625 (2014)
15. Zhang, D., Lu, G.: Review of shape representation and description techniques. PR **37**(1), 1–19 (2004)