



Evaluation Framework of Superpixel Methods with a Global Regularity Measure

Rémi Giraud, Vinh-Thong Ta, Nicolas Papadakis

► To cite this version:

Rémi Giraud, Vinh-Thong Ta, Nicolas Papadakis. Evaluation Framework of Superpixel Methods with a Global Regularity Measure. Journal of Electronic Imaging, 2017, Special Section on Superpixels for Image Processing and Computer Vision. hal-01519635

HAL Id: hal-01519635

<https://hal.science/hal-01519635>

Submitted on 9 Jun 2017

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.

Evaluation Framework of Superpixel Methods with a Global Regularity Measure

Rémi Giraud^{a,b,c,d,*}, Vinh-Thong Ta^{a,b,e}, Nicolas Papadakis^{c,d}

^aUniversity of Bordeaux, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.

^bCNRS, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.

^cUniversity of Bordeaux, IMB, UMR 5251, F-33400 Talence, France.

^dCNRS, IMB, UMR 5251, F-33400 Talence, France.

^eBordeaux INP, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.

Abstract.

In the superpixel literature, the comparison of state-of-the-art methods can be biased by the non-robustness of some metrics to decomposition aspects, such as the superpixel scale. Moreover, most recent decomposition methods allow to set a shape regularity parameter, which can have a substantial impact on the measured performances. In this paper, we introduce an evaluation framework, that aims to unify the comparison process of superpixel methods. We investigate the limitations of existing metrics, and propose to evaluate each of the three core decomposition aspects: color homogeneity, respect of image objects and shape regularity. To measure the regularity aspect, we propose a new global regularity measure (GR), which addresses the non-robustness of state-of-the-art metrics. We evaluate recent superpixel methods with these criteria, at several superpixel scales and regularity levels. The proposed framework reduces the bias in the comparison process of state-of-the-art superpixel methods. Finally, we demonstrate that the proposed GR measure is correlated with the performances of various applications.

Keywords: Superpixels, Evaluation Framework, Superpixel Metrics, Regularity.

*Rémi Giraud, remi.giraud@labri.fr

1 Introduction

Superpixel decomposition methods, that group pixels into homogeneous regions, have become popular with Ref. 1, and have been widely proposed in the past years.^{2–16} Such decompositions provide a low-level image representation, while trying to respect the image contours. The superpixels are usually used as a pre-processing in many computer vision methods such as: object localization,¹⁷ stereo and occlusion processing,¹⁸ contour detection and segmentation,¹⁹ multi-class object segmentation,^{20–24} data associations across views,²⁵ face labeling,²⁶ or adapted neural networks architectures.^{27,28} For most methods, the computational cost depends on the number of elements to process, and the superpixel representation is well adapted since this aspect is drastically reduced. Moreover, superpixels can also directly improve the accuracy of applications such as labeling,¹⁹

since they gather homogeneous pixels in larger areas, thus reducing the potential noise of independent pixel labeling. Finally, contrary to the classical multi-resolution approach, that decreases the image size by averaging the information, the superpixels preserve the image geometry and content.

The definition of an optimal superpixel decomposition depends on the application. Nevertheless, according to the literature, the desired properties of a superpixel method should be the following. (i) The color clustering must group pixels into homogeneous areas in terms of color features, for instance in RGB or CIELab space. (ii) The decomposition boundaries should adhere to the image contours, *i.e.*, the superpixels should not overlap with several image objects. (iii) The superpixels should have regular shapes and consistent sizes within the decomposition. Regular superpixels provide an easier analysis of the image content, and reach better performances for applications such as tracking.^{29,30} Since these properties cannot be optimal at the same time,³¹ most decomposition methods compute a trade-off between these aspects in their model. For instance, Figure 1 shows synthetic examples (a) and (d) where a trade-off between the considered aspects must be computed to decompose the images into three superpixels. In Figures 1(b) and (e), the color homogeneity (i) is optimal, while the respect of image contours (ii) and shape regularity (iii) are respectively favored in (c) and (f). Such examples enable to illustrate how a superpixel decomposition method should behave to optimize each criteria.

Many evaluation metrics dedicated to the superpixel context were proposed to evaluate the consistency of a decomposition with respect to the three above-mentioned superpixel properties.^{5,7,8,32–34} The metrics namely include intra-cluster variation (ICV)³⁵ or explained variation (EV)⁴ for color homogeneity (i), the undersegmentation error (UE),^{5,8,33} achievable segmentation accuracy (ASA),⁷ or boundary recall (BR)³⁶ for the criteria (ii), on the respect of image objects, and circularity (C)³² or mismatch factor¹² for the criteria (iii), of superpixel shape regularity. These

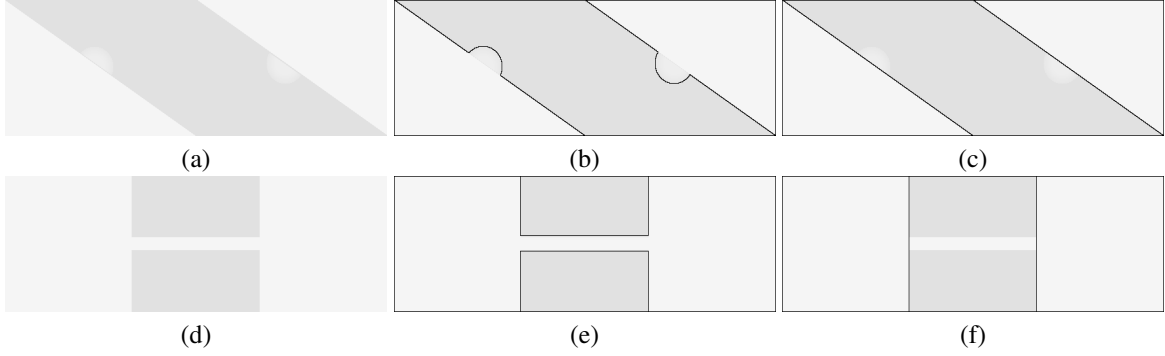


Fig 1 Examples of trade-off between the superpixel properties for the decomposition of images (a) and (d) into three superpixels. The decomposition in (b) and (e), are optimal in terms of color homogeneity while the ones in (c) and (f) are respectively optimal in terms of respect of image contours and regularity.

metrics offer different interpretations and evaluations of the superpixel properties. For instance, the circularity³² and the local regularity metric introduced in Ref. 34 respectively evaluate circular and square shapes as the most regular ones, which can be arguable.

Many recent methods allow the user to set a parameter that enforces or relaxes the regularity constraint of the superpixel shape, *e.g.*, Ref. 8, 11–14, 16. We show in Figure 2 that a same superpixel method can produce different results, according to this parameter. While regular decompositions provide visually consistent superpixels, with approximately the same shape and size, the irregular one more accurately follows the color variations, at the expense of dissimilar superpixel shapes.

Since the regularity can be set in most recent methods,^{8,9,11–14,16} this parameter should be carefully tuned during the evaluation of performances. In each superpixel paper that introduces a decomposition method with such parameter, the authors give a default regularity setting that generally provides a trade-off between segmentation accuracy and superpixel shape regularity. However, when comparing the state-of-the-art methods, even if the results are computed with the default settings of the initial paper, there can be an important bias in the comparison, since most evalua-

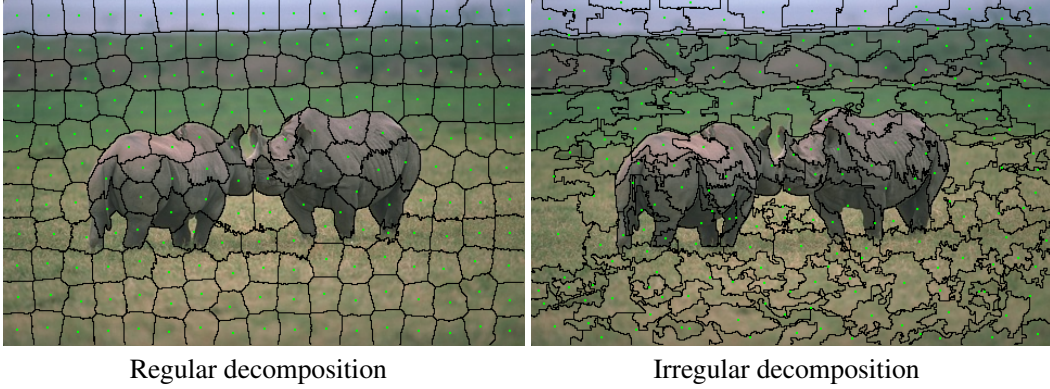


Fig 2 Example of decompositions obtained with the same method⁸ with different regularity settings. The difference between the decompositions can be expressed with the variance of distances between superpixel barycenters (normalized by the average distance), which is 0.632 for the irregular decomposition and 0.406 for the regular one.

tion metrics or superpixel-based pipelines achieve very different performances depending on the regularity of the decompositions.

In the recent study of Ref. 37, no mention is made on the used settings, although the ranking between methods could be altered with different regularity parameters, while Ref. 38 shows that the regularity can be in line with the performances of segmentation and labeling applications. Therefore, we recommend to evaluate the performances with accurate metrics, and for several superpixel scales and regularity levels. Hence, the performances of superpixel methods are not biased by the chosen regularity setting.

1.1 Contributions

In this paper, we propose an evaluation framework of superpixel methods. This work aims to unify the evaluation process across superpixel works and to enable a clear assessment of their performances.

- We take a global view of existing superpixel metrics and investigate their limitations. We show that the evaluation can be reduced to one criteria per decomposition property: color

homogeneity, respect of image objects and shape regularity.

- We address the non-robustness of the state-of-the-art regularity metrics, with the proposed global regularity (GR) measure, that relevantly evaluates the shape regularity and consistency of the superpixels.¹
- We report an evaluation of the state-of-the-art methods on this GR criteria. Contrary to the standard superpixel literature, to reduce the bias in the evaluation process, we recommend to perform the evaluation for several regularity levels, evaluated with GR, since it expresses the range of potential results of each superpixel method.
- Finally, we demonstrate on various applications that the regularity constraint has a substantial impact on performances, and that our GR metric is correlated with the obtained results.

1.2 Outline

In this paper, we first present in Section 2 the existing superpixel metrics, and investigate their limitations. To address the non-robustness of the regularity metrics, we propose in Section 3 a new global regularity measure. In Section 4, we use the considered metrics to compare the state-of-the-art superpixel methods according to several superpixel scales and regularity levels. Finally, we demonstrate the impact of the regularity setting on several applications in Section 5.

2 Standard Superpixel Metrics and Limitations

In this section, we present the existing superpixel metrics, that were progressively introduced in the literature to evaluate the color homogeneity, the respect of image objects or the shape regularity, and we show their limitations to only focus on relevant ones. In Sections 2.1 and 2.3, we present

¹An implementation of the proposed GR metric is available at: www.labri.fr/~rgiraud/downloads

color homogeneity and shape regularity metrics that evaluate for an image I , a superpixel decomposition $\mathcal{S} = \{S_k\}_{k \in \{1, \dots, |\mathcal{S}|\}}$ composed of $|\mathcal{S}|$ superpixels S_k , where $|\cdot|$ denotes the cardinality of the considered element. In Section 2.2, we present metrics that evaluate the respect of image objects and compare the decomposition to a ground truth denoted $\mathcal{G} = \{G_j\}_{j \in \{1, \dots, |\mathcal{G}|\}}$, with G_j a segmented region.

2.1 Homogeneity of Color Clustering

The homogeneity of the color clustering is a core aspect of the superpixel decomposition. Most methods compute a trade-off between spatial and color distances to compute the superpixels. The ability to gather homogeneous pixels should hence be considered in the comparison process, but this aspect is rarely evaluated in state-of-the-art superpixel works. In Section 5, we show that color homogeneity is particularly interesting for image compression.^{5,39}

The homogeneity of colors can be evaluated by comparing the average colors of superpixels to the colors of the pixels in the initial image. The intra-cluster variation (ICV)³⁵ has been proposed to measure such color variation for a decomposition \mathcal{S} :

$$\text{ICV}(\mathcal{S}) = \frac{1}{|\mathcal{S}|} \sum_{S_k} \frac{\sqrt{\sum_{p \in S_k} (I(p) - \mu(S_k))^2}}{|S_k|}, \quad (1)$$

with $\mu(S_k)$, the average color of the superpixel S_k .

The explained variation (EV)⁴ was also proposed to evaluate the homogeneity of the color clustering and is defined as:

$$\text{EV}(\mathcal{S}) = \frac{\sum_{S_k} |S_k| (\mu(S_k) - \mu(I))^2}{\sum_{p \in I} (I(p) - \mu(I))^2}. \quad (2)$$

Limitations

The ICV metric presents several drawbacks. It is not normalized by the image variance to allow comparable evaluation between different images,³⁷ and the superpixel size is not considered, so the measure is not robust to the superpixel scale. In Figure 3, we illustrate these issues on synthetic examples. The dynamic and size of the same image with the same decomposition are altered, and these transformations impact the measure of ICV.

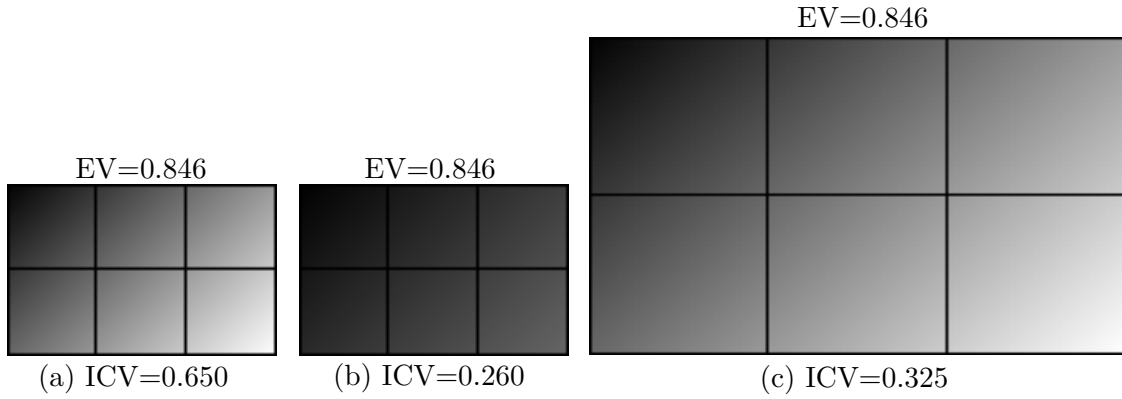


Fig 3 Comparison of ICV and EV to measure the superpixel color homogeneity. The dynamic and dimension of the same image (a) are respectively modified in (b) and (c). EV is robust to these transformations contrary to ICV.

The EV metric is robust to these transformations. It expresses the color variance contained into each superpixel, and simple calculations provide another explicit formulation:

$$\text{EV}(\mathcal{S}) = 1 - \sum_{S_k} \frac{|S_k|}{|I|} \cdot \frac{\sigma(S_k)^2}{\sigma(I)^2}. \quad (3)$$

In Figure 4(c), we illustrate the color variance $\sigma(S_k)^2$ within each superpixel S_k . Higher variance within a superpixel will lower the EV criteria, so high EV values express homogeneous color clustering. Despite the robustness of this criteria, it was not considered in the main state-of-the-art superpixel works. As in Ref. 37, we recommend the use of EV to evaluate color homogeneity.

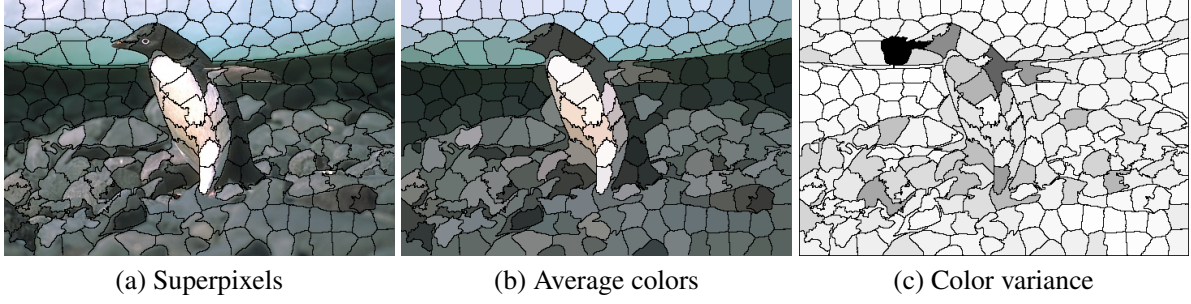


Fig 4 Example of superpixel decomposition (a), with average colors (b), and color variance $\sigma(S_k)^2$ within each superpixel S_k (c).

2.2 Respect of Image Objects

The most considered aspect in the superpixel evaluation is the respect of the image objects. An accurate decomposition should have superpixels that do not overlap with multiple objects. To evaluate this aspect, many metrics were proposed. In the literature, undersegmentation error (UE), achievable segmentation accuracy (ASA) and boundary recall (BR) measures are mainly reported. While UE and ASA compute overlaps with the ground truth regions, BR is a contour adherence measure, which is usually correlated to the first two metrics.

Regarding the UE, several definitions were proposed.^{5,8,33} UE evaluates the number of pixels that cross ground truth region boundaries. The initial UE formulation,⁵ denoted UE_L , is defined as:

$$UE_L(\mathcal{S}, \mathcal{G}) = \frac{1}{|\mathcal{G}|} \sum_{G_j} \frac{\left(\sum_{S_k, S_k \cap G_j \neq \emptyset} |S_k| \right) - |G_j|}{|G_j|}. \quad (4)$$

The UE_L measure was discussed in several works, *e.g.*, Ref. 8,9,33, since any superpixel that has an overlap with the ground truth segment penalizes the metric. Therefore, UE_L is very sensitive to small overlaps and does not accurately reflect the respect of the image objects. A method to reduce the overlap considered in (4) was proposed in Ref. 8, but it requires a parameter. Recent

state-of-the-art works tend to use the free parameter formulation of UE introduced by Ref. 33:

$$\text{UE}(\mathcal{S}, \mathcal{G}) = \frac{1}{|I|} \sum_{S_k} \sum_{G_j} \min\{|S_k \cap G_j|, |S_k \setminus G_j|\}. \quad (5)$$

This formulation respectively considers the intersection or the non-overlapping superpixel area in case of small or large overlap with the ground truth region, and addresses the non-robustness of former UE definitions.

The ASA measure⁷ also aims at evaluating the overlap of superpixels with the ground truth. It is reported in most of the superpixel literature as follows:

$$\text{ASA}(\mathcal{S}, \mathcal{G}) = \frac{1}{|I|} \sum_{S_k} \max_{G_j} |S_k \cap G_j|. \quad (6)$$

For each superpixel S_k , the largest possible overlap with a ground truth region G_j is considered, and higher values of ASA indicate better results.

Another measure is extensively reported in the superpixel literature to evaluate the adherence to object contours: the boundary recall (BR). This metric evaluates the detection of ground truth contours $\mathcal{B}(\mathcal{G})$ by the superpixel boundaries $\mathcal{B}(\mathcal{S})$ such that:

$$\text{BR}(\mathcal{S}, \mathcal{G}) = \frac{1}{|\mathcal{B}(\mathcal{G})|} \sum_{p \in \mathcal{B}(\mathcal{G})} \delta\left[\min_{q \in \mathcal{B}(\mathcal{S})} \|p - q\| < \epsilon\right], \quad (7)$$

with $\delta[a] = 1$ when a is true and 0 otherwise, and ϵ is a distance threshold that has to be set, for instance to 2 pixels.^{7,9} Each ground truth pixel is considered as detected, if a superpixel boundary is at less than an ϵ distance.

Limitations

Although the BR measure (7) has been extensively reported in the literature and is recommended in Ref. 37, it does not express the respect of image objects or the contour detection performances, as it only considers true positive contour detection. Hence, the number of computed superpixel contours is not considered, and very irregular methods can obtain higher BR results. Figure 5 compares two decompositions with the same number of superpixels that have maximal $BR = 1$. One of the decomposition is very irregular and produces a high number of boundary superpixels, and this aspect is not considered in the BR metric. For these reasons, recent works such as Ref. 12, 16, 40 report BR results according to the contour density (CD), which measures the number of superpixel boundaries such that $CD(\mathcal{S}) = |\mathcal{B}(\mathcal{S})|/|I|$. Nevertheless, BR needs a parameter ϵ to be set and is not sufficient to reflect the respect of image objects. BR should only be considered as a tool for evaluation of contour detection performances, as shown in Section 5.

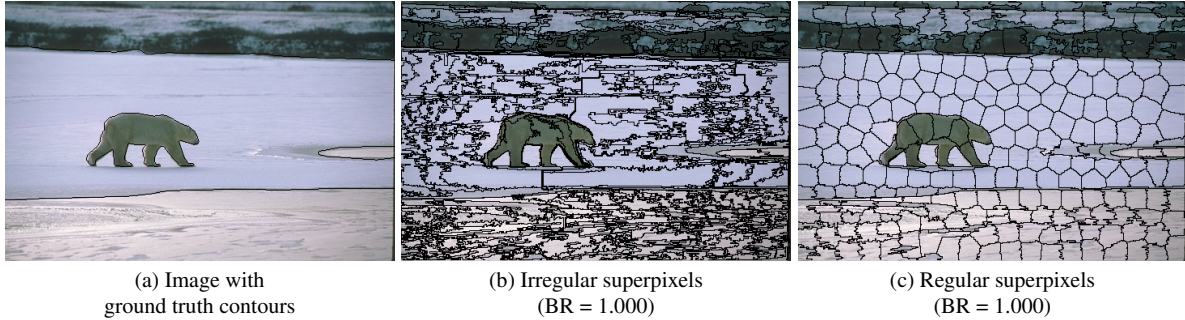


Fig 5 Examples of irregular¹⁴ (b) and regular⁸ (c) decomposition of an image (a), with maximal BR measure. BR evaluates the detection of ground truth contours so both decompositions can have maximal BR measure, although the irregular one produces more superpixel boundaries.

The study of Ref. 37 claims that the UE (5) and ASA (6) measures are correlated, so that $ASA = 1 - UE$, making redundant the use of both for superpixel evaluation. Under relevant

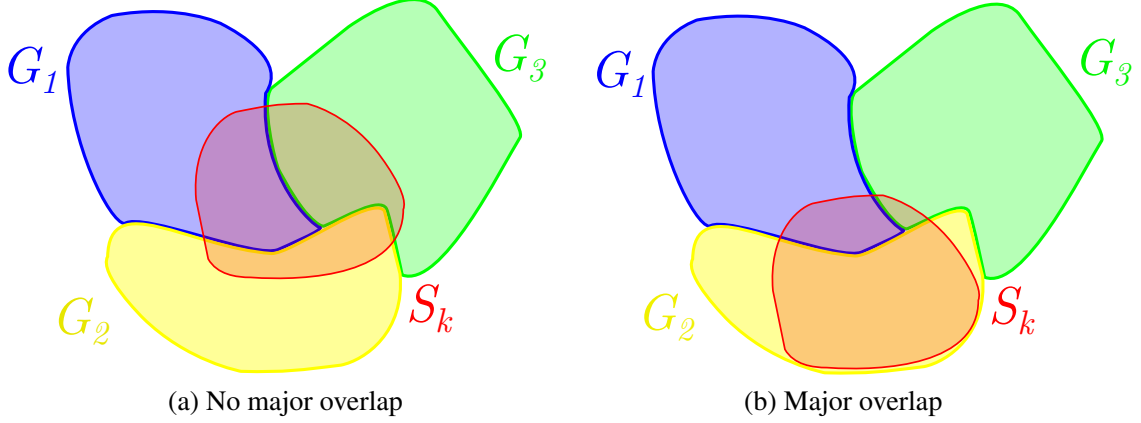


Fig 6 Examples of decomposition where a superpixel S_k overlap with multiple ground truth regions G_j . In (b), contrary to (a), a region (G_2) overlaps with more than half of S_k , which corresponds to the major overlap hypothesis ensuring (8).

assumptions, we now show that the true relation between the two measures is:

$$\text{ASA}(\mathcal{S}, \mathcal{G}) = 1 - \text{UE}(\mathcal{S}, \mathcal{G})/2. \quad (8)$$

In Appendix A, we demonstrate that the relation (8) is exact when all superpixels have a major overlap, *i.e.*, a ground truth region intersects with more than half of the superpixel area. This case is illustrated in Figure 6(b), where a superpixel S_k has a major overlap with a region G_2 . Note that this assumption is necessarily true when each superpixel overlaps with only two regions or when the ground truth is binary. In Appendix A, we measure the error of (8) on state-of-the-art methods. This error appears to be negligible, underlying the likelihood of this assumption.

Hence, since the UE_L measure has been proven to be non-robust, and we demonstrate that the study of UE and ASA is equivalent, we recommend to restrain the evaluation of the respect of image objects to the ASA results (6).

2.3 Regularity

Supapixel decompositions compute a lower-level image representation. Since the desired number of elements is given, and most methods tend to compute regions with approximately the same size, superpixels are generally created in smooth areas (see for instance Figure 2). To produce a consistent decomposition, most superpixel methods compute a trade-off between color clustering and regularity, which has been proven to have an impact on application performances.^{6,30,38} Therefore, the regularity of the superpixel shapes is a core aspect when evaluating and comparing decomposition methods.

In Ref. 32, the circularity metric was introduced to locally evaluate the compactness of the superpixels. This measure is the usual local regularity metric, and has been considered in state-of-the-art works,^{11,16,30,40} and benchmarks.^{34,37,41} This circularity C is defined for a superpixel shape S as follows:

$$C(S) = \frac{4\pi|S|}{|P(S)|^2}, \quad (9)$$

where $P(S)$ is the superpixel perimeter. The regularity is hence considered as the ability to produce circular areas.

While the circularity independently evaluates the local compactness of each superpixel, other works propose to evaluate the consistency of the shapes within the decomposition. In Ref. 34, the variance of superpixel sizes is considered. Ref. 12 goes further and proposes an adapted version of the mismatch factor⁴² to measure the consistency in terms of size and shape. The mismatch factor is computed as $1 - J(S_1, S_2)$ with J being the standard Jaccard index,⁴³ that computes the overlap between two regions S_1 and S_2 such that: $J(S_1, S_2) = |S_1 \cap S_2| / |S_1 \cup S_2|$. The global measure J

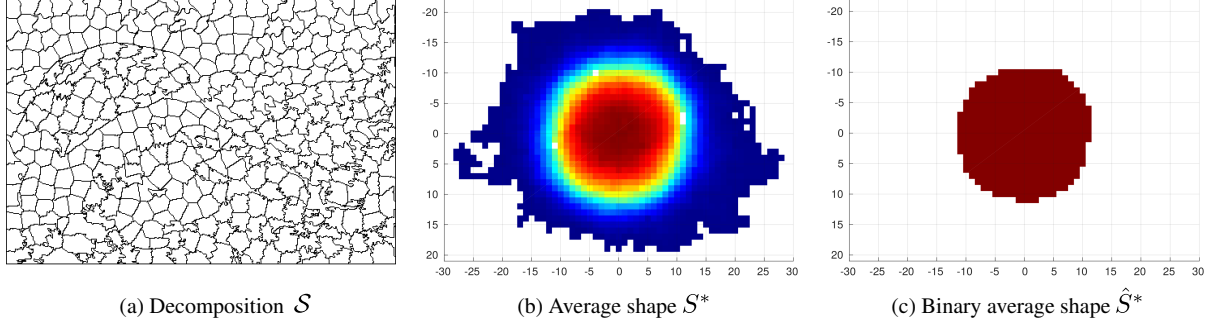


Fig 7 Illustration of the average superpixel shape definition.¹² A decomposition is considered in (a). The superpixel shapes are registered and averaged into S^* (b) to provide the binary average shape (c).

for a superpixel decomposition \mathcal{S} is computed as follows:¹²

$$J(\mathcal{S}) = \frac{1}{|\mathcal{S}|} \sum_{S_k} J(S_k^*, \hat{S}^*), \quad (10)$$

with S_k^* the registered superpixel S_k , so its barycenter is at the origin of the coordinates system, and \hat{S}^* , the binary average shape of \mathcal{S} . In the following, we consider the J metric when comparing to Ref. 12, since, contrary to the mismatch factor, its interpretation is consistent with the other presented metrics, *i.e.*, higher value is better. Eq. (10) compares each superpixel to the binary average shape \hat{S}^* of the decomposition using J . To compute \hat{S}^* , the registered superpixels S_k^* are first averaged into S^* such that:

$$S^* = \frac{1}{|\mathcal{S}|} \sum_{S_k} S_k^*. \quad (11)$$

A thresholded shape S_t^* is then defined by binarization with respect to a threshold t . The binary average shape is finally defined as $\hat{S}^* = S_{\underset{t}{\operatorname{argmax}}(|S_t^*| \geq \mu)}^*$, with $\mu = |I|/|\mathcal{S}|$, the average superpixel size. We illustrate these definitions in Figure 7. We represent a superpixel decomposition \mathcal{S} , the corresponding average of superpixel shapes S^* , and the binary average shape \hat{S}^* .

Limitations

Although the circularity metric (9), introduced in Ref. 32, is the local regularity metric usually used in the literature,^{11,16,30,37,40,41} it presents several drawbacks. For instance, the relevance of this metric, and the regularity definition have been discussed,¹² since the circularity considers a circle or a hexagon as more regular than a square, and is very sensitive to boundary smoothness.⁴⁴ As a consequence, some methods, such as Ref. 40, start from a hexagonal repartition of superpixels seeds, and design their spatial constraint to fit to a hexagonal grid in order to obtain higher circularity. However, the superpixel literature more generally refers to the regularity as the ability to produce convex shapes with smooth boundaries. Therefore, squares, circles or hexagons should be considered as regular shapes. Moreover, since most superpixel methods generate a regular square grid with their regularity parameter set to maximum value, it would make sense to obtain the highest regularity measure for such square decomposition.

The mismatch factor or Jaccard index (10) was introduced by Ref. 12 to evaluate the shape consistency within the whole decomposition, but it does not consider the superpixel size, so each shape equally contributes. Moreover, by computing a thresholded average shape, the metric appears to be non-robust to large shape outliers, leading to potential irrelevant measures. Finally, with such metric that only considers the shape consistency, decomposing an image with its lines or with stretched rectangles would give the highest regularity measure.

Contrary to ASA and EV, that give a relevant measure of aspects (i) color homogeneity and (ii) respect of image objects, no existing metric provides a robust and accurate regularity measure of a superpixel decomposition. As a consequence, we propose in the following Section 3, a new global regularity measure, that addresses the limitations of state-of-the-art ones.

3 A new Regularity Measure

In the literature, the circularity (9) has been proposed to measure the shape regularity,³² and the mismatch factor to evaluate the shape consistency across the decomposition.¹² These measures present several drawbacks, that we address in this work by introducing two new metrics (14) and (15), combined into the proposed global regularity (GR) measure. GR robustly evaluates both shape regularity and consistency over the decomposition, and we demonstrate in Sections 4 and 5 that it relevantly evaluates the performances of superpixel methods.

3.1 Shape Regularity

As stated in Section 2.3, an accurate superpixel shape regularity measure should provide the highest results for convex shapes, such as squares, circles or hexagons, and penalize unbalanced shapes while considering noisy boundaries. To express such a measure, we propose to combine all these aspects into a new shape regularity criteria (SRC).⁴⁴ The convexity of a shape S is considered, *i.e.*, the smoothness of its contours and the overlap with its convex hull H_S , that entirely contains S . To evaluate both overlap with the convex hull and contour smoothness, we first define as $CC(S) = |P(S)|/|S|$, the ratio between the perimeter and the area of a shape S , which is linked with the Cheeger constant for convex shapes.⁴⁵ Then, we introduce our criteria of regularity (CR) as:

$$CR(S) = \frac{CC(H_s)}{CC(S)}. \quad (12)$$

Since the convex hull H_s entirely contains S and has a lower perimeter, the CR measure is between 0 and 1, and is maximal for convex shapes such as squares, circles or hexagons. Nevertheless, the comparison to the convex hull is not sufficient to define the regularity of a superpixel.

The balanced repartition of the pixels within the shape is another aspect to consider. Otherwise, shapes such as ellipses or lines would get the maximum CR results. We define the variance term V_{xy} as the ratio between the minimum and maximum variance of pixel positions x and y that belong to S :

$$V_{xy}(S) = \frac{\min(\sigma_x, \sigma_y)}{\max(\sigma_x, \sigma_y)}, \quad (13)$$

with σ_x and σ_y , the standard deviations of the pixel positions. Such measure enables to penalize dissimilarity in the pixel repartition, and $V_{xy} = 1$ if, and only if, $\sigma_x = \sigma_y$, *i.e.*, if the spatial pixel repartition is balanced.

The proposed shape regularity criteria (SRC) is defined as follows:

$$SRC(\mathcal{S}) = \sum_{S_k} \frac{|S_k|}{|I|} CR(S_k) V_{xy}(S_k). \quad (14)$$

Note that in practice, we use the square root of V_{xy} , so both criteria have similar variation ranges. SRC robustly and jointly evaluates convexity, contour smoothness and balanced pixel repartition.

Circularity vs Shape Regularity Criteria

To demonstrate the robustness and relevance of SRC (14) over circularity C (9), we consider in Figure 8 synthetic shapes that are split into three groups (regular, standard and irregular), and generated with smooth (top) and noisy boundaries (bottom). First, we present the circularity drawbacks, which reports lower measure for the Square than for the Hexagon, or the Ellipse. Since methods such as Ref. 12, 40 produce superpixels from a hexagonal grid, the regularity evaluation is very likely to be superior for these methods than for other ones starting from square regular grids. The circularity is also very sensitive to the contour smoothness, so regular and standard



















	Regular shapes			Standard shapes			Irregular shapes		
	Square	Circle	Hexagon	Ellipse	Cross	Bean	W	Split	U
C									
	0.830	1.000	0.940	0.870	0.530	0.580	0.150	0.280	0.150
SRC	1.000	0.989	0.987	0.712	0.650	0.564	0.387	0.369	0.233
C									
	0.480	0.430	0.420	0.410	0.340	0.440	0.100	0.210	0.070
SRC	0.716	0.633	0.625	0.474	0.515	0.500	0.296	0.321	0.136

Fig 8 Comparison of circularity (C) and proposed shape regularity criteria (SRC) on three groups of synthetic shapes with smooth (top) and noisy boundaries (bottom). C appears to only favor circular shapes and does not separate standard and regular noisy ones. The SRC metric addresses these issues and clearly differentiates the shape groups in the smooth and noisy cases.

noisy shapes have similar measure, and the groups are no longer differentiated (see the bottom part of Figure 8). Finally, standard but smooth shapes have much higher circularity than noisy regular ones.

With SRC, we first note that the three regular shapes have the highest regularity measure (≈ 1), and that regular but noisy shapes, have similar SRC to the smooth standard ones. Overall, since SRC is less dependent on the boundary smoothness, in each group, smooth and noisy shapes are clearly separated, contrary to C. Finally, we show the metric evolution with the shape size in Figure 9. As stated in Ref. 46, due to discretization issues, the circularity must be thresholded so it is not superior to 1, and it drops as the shape size increases. Therefore, this metric is not robust to the superpixel size, and the comparison of methods on circularity is relevant only if the compared superpixel decompositions have the same number of elements. Contrary to the circularity, the SRC metric is robust to the superpixel scale and provides a consistent evaluation of shape regularity.

Nevertheless, the SRC and circularity metrics independently evaluate each superpixel shape

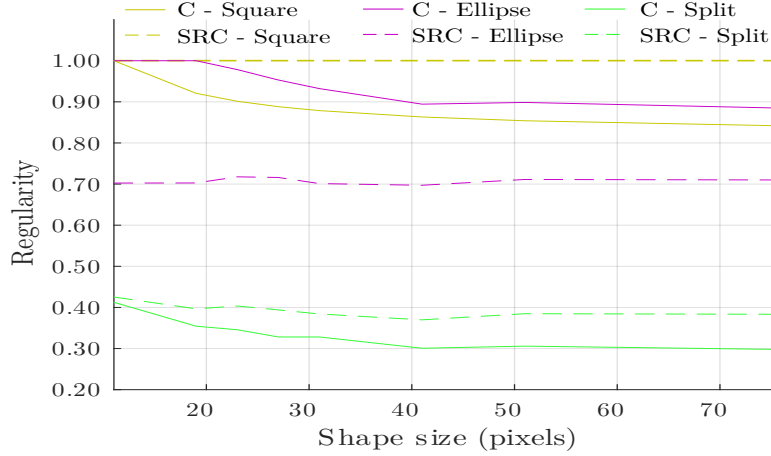


Fig 9 Robustness to the superpixel scale of the proposed shape regularity criteria (SRC) compared to the circularity (C).

without considering the global homogeneity of the decomposition. Superpixel methods, *e.g.*, Ref. 2, 15, or other segmentation algorithms such as the quadtree partitioning method⁴⁷ can produce regions of very variable sizes. In Figure 10, we show an example of decomposition with superpixels having approximately the same size,⁸ and a standard quadtree-based partition,⁴⁷ which produces larger regions in areas with lower color variance. Since the circularity and SRC measures independently consider each superpixel and report an average evaluation of local regularity, the quadtree partition obtains the highest measure although its elements do not have similar sizes. Such local metrics are thus not sufficient to express the global regularity of a decomposition.

3.2 Shape Consistency

The regularity of the superpixel shapes evaluated with a local criteria is a relevant information but does not reflect their consistency in terms of shape and size within the decomposition. In Section 3.1, we define the shape regularity properties for a superpixel: convexity, boundary smoothness and balanced pixel repartition. Nevertheless, a perfectly regular decomposition of an image should be composed of similar regular shapes at the same scale. In other words, a relevant criteria should

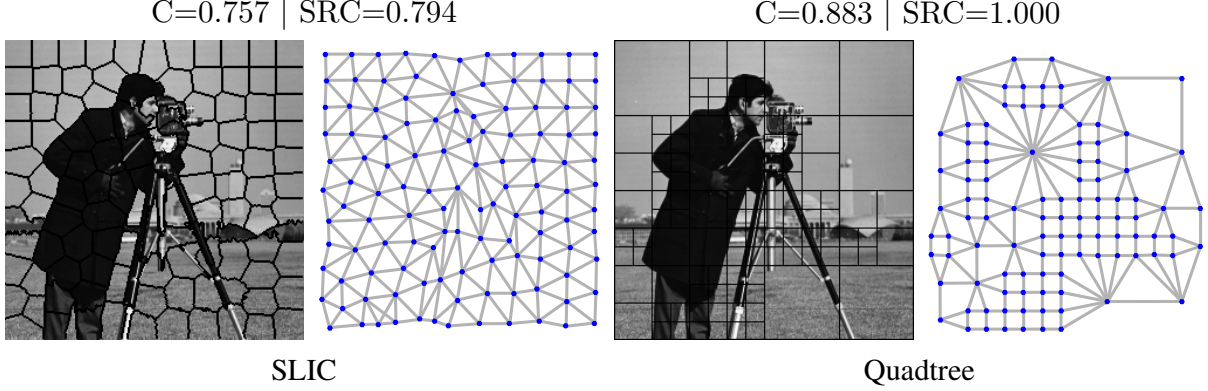


Fig 10 Limitation of the local shape regularity metrics. The shape consistency is not considered so the quadtree decomposition gives higher measures than the decomposition obtained with the SLIC method.⁸ The decompositions are represented with their Delaunay graphs, connecting the barycenters of adjacent superpixels.

locally evaluate the shape regularity and how the superpixels are consistent in terms of shape and size within the decomposition.

As stated in Section 2.3, the mismatch factor (10) uses the Jaccard index to compare the superpixels to a thresholded average shape. As illustrated in Figure 11, this measure can be incoherent with the visual consistency of a decomposition. In these examples, the binary average shape \hat{S}^* is the same, and corresponds to the red areas. The J measure is low and incoherently decreases as the consistency is visually improved.

In this work, we propose the smooth matching factor (SMF), that directly compares the superpixels to the average shape S^* :

$$\text{SMF}(\mathcal{S}) = 1 - \sum_{S_k} \frac{|S_k|}{|I|} \cdot \left\| \frac{S^*}{|S^*|} - \frac{S_k^*}{|S_k^*|} \right\|_1 / 2. \quad (15)$$

SMF compares the spatial distributions of the average superpixel shape S^* to each registered superpixel shape S_k^* . The SMF criteria should be close to 1 if the distributions of pixels within the shapes are similar, and close to 0 otherwise. Overall, the proposed SMF metric evaluates the con-

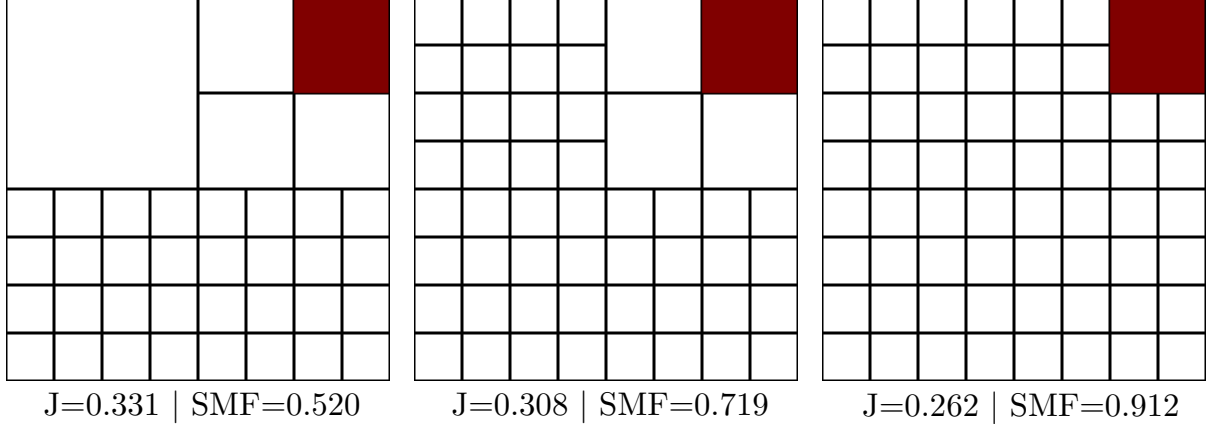


Fig 11 Illustration of several decomposition examples with the corresponding average binary shape in red, and J (10) and SMF (15) values.



Fig 12 Decomposition examples with corresponding J (10) and SMF (15) measures.

sistency in terms of shape and size within a decomposition. By considering the average shape S^* without thresholding, the evaluation is more robust to shape outliers. This criteria addresses the non-robustness of the mismatch factor (J),¹² as can be seen in Figure 11, where SMF is relevant according to the consistency of the decomposition. Finally, examples of superpixel decomposition on natural images are given in Figure 12 and illustrate that SMF is not correlated to the J measure.

3.3 Global Regularity Measure

As previously stated, a perfectly regular superpixel decomposition should be composed of compact shapes, that would be consistent in terms of size and shape. The SRC metric enables to locally

evaluate the shape regularity while the SMF measures the shape consistency. To evaluate both aspects, we propose to combine these metrics in the global regularity (GR) measure:

$$\text{GR}(\mathcal{S}) = \text{SRC}(\mathcal{S})\text{SMF}(\mathcal{S}). \quad (16)$$

In the following Section 4, we report the evaluation on the considered metrics (EV, ASA, GR) for state-of-the-art superpixel methods. These measures are reported according to the number of generated superpixels, and also according to the regularity, since this setting substantially impacts performances.

4 Comparison of State-of-the-Art Superpixel Methods

4.1 Dataset

We compare the performances of state-of-the-art methods on the standard Berkeley segmentation dataset (BSD).⁴⁸ This dataset contains 200 various test images of size 321×481 pixels. For each image, human segmentations are provided and considered as ground truth to evaluate the respect of image objects. At least five decompositions are provided per image, and the presented results in the following sections are averaged on all ground truths. Note that other datasets can be considered, *e.g.*, Ref. 49, 50, but the BSD⁴⁸ is the most used dataset for comparing superpixel methods. Moreover, Ref. 37 shows that decomposition algorithms that perform well on the BSD usually perform well on other datasets.

4.2 Considered Superpixel Methods

Many frameworks have been proposed to decompose an image into superpixels using either graph-based,^{2,6,7,11} watershed,^{12,31,51} coarse-to-fine^{9,14} or gradient-ascent^{3,5,8,13,16} approaches (see Ref.

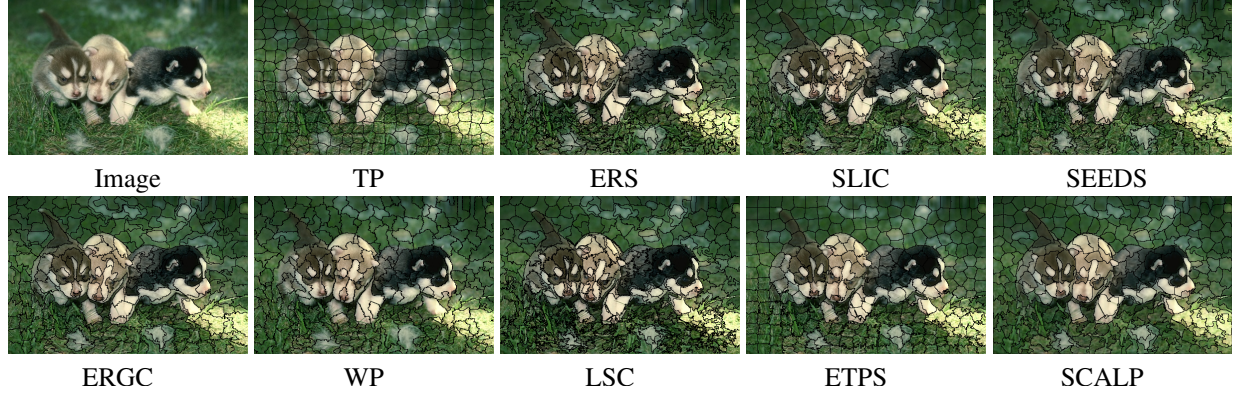


Fig 13 Decomposition example of each considered state-of-the-art superpixel methods for approximately 250 superpixels.

37 for a detailed review of existing methods). In order to illustrate the proposed evaluation framework, and to validate our regularity measure, we consider the following state-of-the-art superpixel methods: TP,⁵ ERS,⁷ SLIC,⁸ SEEDS,⁹ ERGC,¹¹ WP,¹² LSC,¹³ ETPS¹⁴ and SCALP.¹⁶ As stated in the introduction, only the most recent methods, SLIC,⁸ ERGC,¹¹ WP,¹² LSC,¹³ ETPS¹⁴ and SCALP,¹⁶ enable to set a regularity parameter. A decomposition example for each considered method is given in Figure 13.

4.3 Quantitative Results

In this section, we compare the methods considered in Section 4.2 on the recommended metrics for several superpixel scales and several regularity levels. In Figure 14, we perform the standard evaluation of performances, according to the number of generated superpixels. The behavior of each method regarding the different decomposition aspects, homogeneity of color clustering (EV), respect of image objects (ASA), and regularity (GR), is respectively evaluated in Figures 14(a), (b) and (c). Methods such as TP⁵ and WP,¹² that produce very regular superpixels appear to poorly perform on other metrics. Although they report high regularity, recent methods SCALP¹⁶ and ETPS¹⁴ perform well on color homogeneity, evaluated with EV. SCALP¹⁶ even performs best

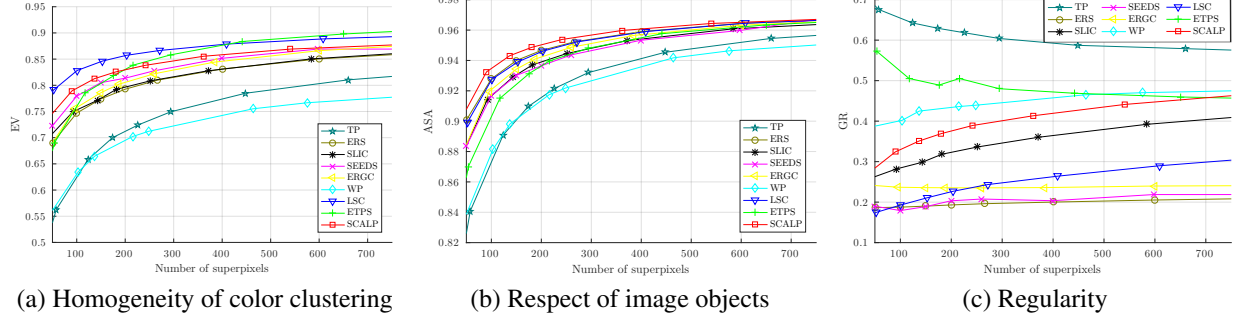


Fig 14 Evaluation of state-of-the-art superpixel methods on EV, ASA and GR according to the number of superpixels, with the methods default regularity settings. The hierarchy between the performances of methods tends to be consistent on all metrics.

on respect of image objects, evaluated with ASA. We summarize this evaluation in Table 1. A superpixel decomposition cannot obtain maximum values on each criteria at the same time, and according to the desired decomposition aspect, one would grant more consideration to a particular criteria.

Table 1 Average EV, ASA and GR on several scales $K = [25, 1000]$, with the methods default regularity settings.

	TP ⁵	ERS ⁷	SLIC ⁸	SEEDS ⁹	ERGC ¹¹	WP ¹²	LSC ¹³	ETPS ¹⁴	SCALP ¹⁶
EV (2)	0.734	0.807	0.809	0.829	0.820	0.713	0.863	0.845	0.840
ASA (6)	0.925	0.951	0.944	0.943	0.947	0.922	0.950	0.947	0.956
GR (16)	0.610	0.197	0.343	0.203	0.237	0.444	0.244	0.487	0.395

Nevertheless, such evaluation is limited, since the hierarchy between methods performances tends to be consistent, *i.e.*, there are no fluctuation in the ranking of methods over a certain number of generated superpixels. Moreover, as stated in the introduction, the setting of regularity has a crucial impact on performances. Therefore, the evaluation should be performed on this criteria, to measure the ability of each method to produce the best trade-off between all evaluation aspects. In Figure 15, we fix the superpixel scale to $K = 250$, and we generate the superpixel results for several regularity levels. The number of superpixels must be carefully set, since some methods may

produce less superpixels than the required number. Larger markers correspond to results obtained with default regularity settings. Hence, for TP,⁵ ERS⁷ and SEEDS⁹ methods, only one value is reported since these methods do not enable to set the regularity of the decomposition.

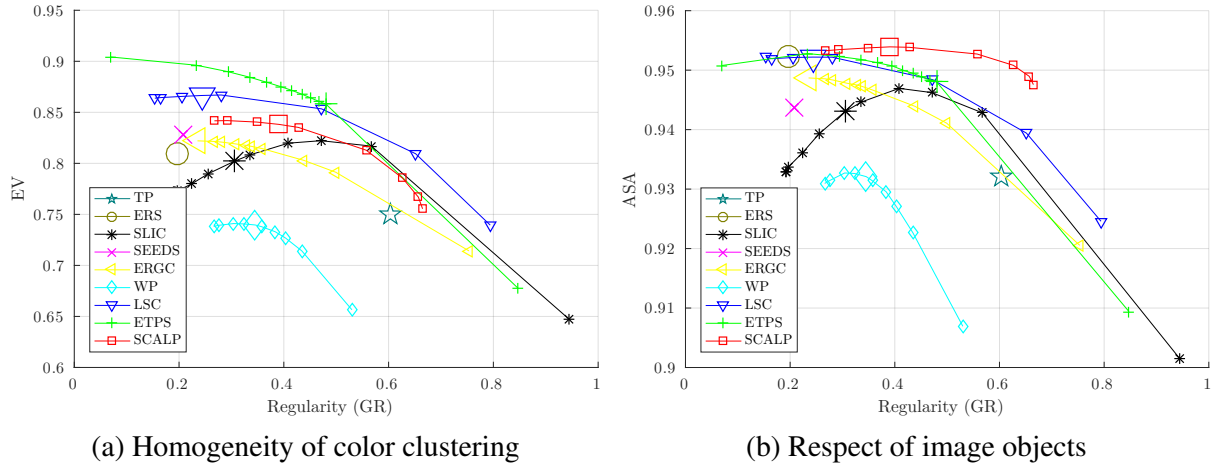


Fig 15 Evaluation of superpixel state-of-the-art superpixel methods on EV, ASA according to the regularity (GR) for $K = 250$.

First, the results appear to be greatly impacted by the regularity setting, so for each method, a wide range of performances can be obtained. Second, contrary to the evaluation according to the number of superpixels, the performance curves of methods cross each others, and the hierarchy between methods can be different according to the regularity parameter. Third, the default setting of most methods is close from a good trade-off between all decomposition aspects but is not necessarily optimal. For instance, ETPS¹⁴ default setting sets high regularity and is not optimal for EV and ASA. The performances of SLIC⁸ can also be improved on all aspects for a higher regularity setting. We can note that contrary to other methods, results on EV and ASA significantly drop with lower regularity for SLIC. This can be explained by the post-processing step that enforces the pixel connectivity within superpixels.⁸ With relaxed regularity constraint, more pixels are disconnected and less accurately grouped during the post-processing.

This evaluation demonstrates that the regularity setting is a crucial parameter to set for all methods. Hence, a huge bias can appear in the comparison process if this parameter is not optimal, or at least set according to the default setting reported by the authors. This is particularly true since the gap of performances may be very small between methods. By carefully comparing performances on the recommended metrics, and on both produced superpixel number and regularity, we ensure accurate evaluation of superpixel methods. We further illustrate the impact of regularity in the following Section 5, where we measure application performances for several regularity levels.

5 Impact of the Regularity Parameter on Application Performances

In this section, we consider several applications, and demonstrate how the regularity can affect the performances. For most applications and superpixel methods, the performances appear to be highly correlated to the proposed GR metric. For nearest neighbor matching or tracking, more regular decompositions enable to capture more similar patterns between images, while for image compression or contour detection applications, lower regularity constraint may enable to reach higher performances.

5.1 Nearest Neighbor Matching

In this section, we investigate the impact of regularity on the superpixel matching accuracy between images. The correlation between the superpixel shape and parsing performances, *i.e.*, segmentation and labeling, has been demonstrated in Ref. 38, where regular grid-based methods appear to give higher pixel-wise results. We propose to compute superpixel matching between regular and irregular decompositions using the superpixel neighborhood structure introduced in Ref. 24. In Figure 16, two decompositions are computed on the same image using both regular (Figures

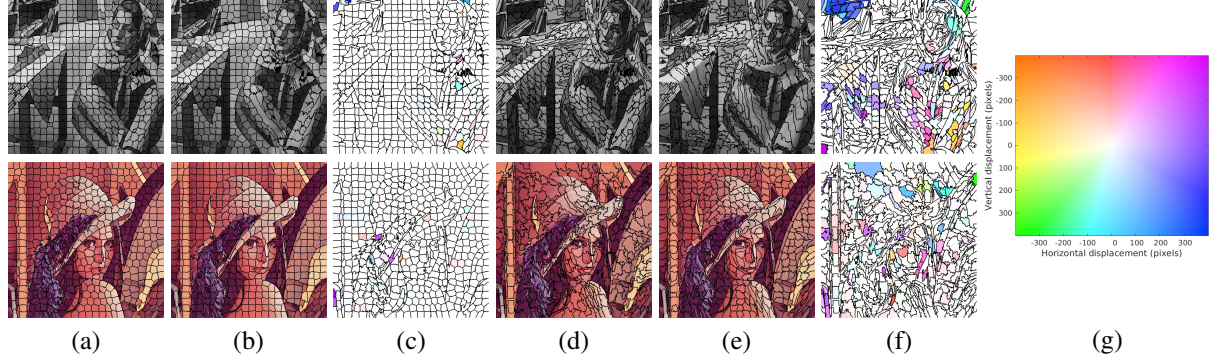


Fig 16 Examples of matching between regular (a), (b) and irregular (d), (e) decompositions. The magnitude of the displacement between the barycenters of the matched superpixels is respectively given for regular and irregular matching in (c) and (f). The displacement is illustrated with the standard optical flow representation (g), and respectively displayed within the superpixel boundaries of images (a) and (d). On these examples, the average GR is respectively 0.478 and 0.205 for the regular (a) and (b) and irregular (d) and (e) decompositions.

16(a) and (b)) and irregular decompositions (Figures 16(d) and (e)). Correspondences are then computed between the two regular decompositions and the two irregular ones, and we compare the respective displacement magnitude of the matches in Figures 16(c) and (f). The correspondences are computed on color features (normalized color histograms), and we display the magnitude of the matches displacement with the standard optical flow representation (Figure 16(g)).

While the correspondences between regular decompositions are accurate (low displacement between matched superpixels), irregular decompositions mislead the matching. With irregular methods, the barycenters may fall outside the superpixels, making difficult the use of the geometrical information of the neighborhood. Regular methods produce decompositions where all elements globally have the same size and the barycenters are usually contained into the superpixels. Therefore, the spatial relation between a superpixel barycenter and the ones of its neighbors is more relevant, which enables more accurate matching.

5.2 Superpixel Tracking

The computational gain obtained with superpixels is particularly interesting in real-time application such as video tracking.^{29,30,52} Two main approaches exist for superpixel-based video segmentation and tracking. The first one computes 3D superpixels, *i.e.*, supervoxels, with the same properties than for 2D methods. Hence, it requires the whole sequence to produce a decomposition and cannot be directly computed with the acquisition of new frames.

The second approach computes superpixels on each frame and the decomposition of the following ones are inferred from the previous segmentations.²⁹ In these methods, optical flow computation is generally used to transfer the superpixels barycenter from a frame to another one.²⁹ As a consequence, a superpixel can only have a unique correspondence in the next frame, and elements describing an object can then be tracked across the video sequence.

We investigate the impact of regularity on the tracking performances of such approach with the temporally consistent superpixels (TSP) method.²⁹ We consider images from Ref. 53, which is composed of short video sequences provided with binary labeling ground truth, and we compare the tracking performances obtained with TSP²⁹ generating regular and irregular decompositions. In Table 2, we report the pixel-wise labeling accuracy over the whole sequences, which expresses the ability of the method to track superpixels. With more regular decompositions, the superpixel shapes and sizes are consistent over time, which enhances the tracking accuracy. We also report the average percentage of new superpixels created between two decompositions. Hence, higher values for irregular decompositions indicate the loss of tracking during the process. Examples of tracking for such regular and irregular decompositions are illustrated in Figure 17 for the considered sequences. The consistency of the regular decompositions enables to more efficiently track

the superpixels, *i.e.*, to compute one-to-one correspondences through the sequences.

Table 2 Tracking accuracy with TSP²⁹ on sequences from Ref. 53. The labeling accuracy and the percentage of lost superpixels over frames are reported for regular and irregular settings.

	Covering accuracy		Lost tracking	
	Regular	Irregular	Regular	Irregular
birdfall2	98.3%	97.8%	1.0%	1.4%
girl	51.1%	50.4%	13.9%	24.8%
parachute	75.3%	73.9%	4.5%	5.1%
penguin	94.3%	85.0%	2.6%	8.8%
average	79.8%	76.7%	5.5%	10.0%



Fig 17 Examples on superpixel tracking performances of TSP²⁹ over the sequences of Ref. 53 for regular and irregular decompositions. On these examples, the average GR is respectively 0.308 and 0.554 for the irregular and regular decompositions.

5.3 Image Compression

Since superpixels define a lower-level representation, they can be used to compress the image content.^{5,39} We propose to study the impact of regularity on the image compression efficiency. Similarly to Ref. 39, each superpixel color channel is approximated by a polynomial of third order.

The color information is then contained into the polynomial coefficients which are computed by a least mean square approach.

A compression example is illustrated in Figure 18 for a regular and an irregular decomposition. We show the reconstructed image I_r from the polynomial coefficients stored for each superpixel. The irregular decomposition produces a reconstructed image closer to the initial one I , while the boundaries of superpixels are visually more noticeable with the regular one. To further evaluate the compression efficiency, we report in Figure 19 the mean square error (MSE) between the initial image and the reconstructed one:

$$\text{MSE}(I, I_r) = \frac{1}{|I|} \sum_{p \in I} (I(p) - I_r(p))^2. \quad (17)$$

The decompositions of state-of-the-art superpixel methods are computed at several regularity levels, for $K = 250$ superpixels, and the MSE (17) results are averaged on all BSD images. For most methods, the default regularity parameter is not optimal to perform well on image compression. These results are in line with the evaluation on the EV metric in Figure 15(a), since homogeneous superpixels in terms of colors enable to compute polynomial coefficients that more accurately fit to the initial image content. However, such decompositions are obtained at the expense of lower regularity.

5.4 Contour Detection and Segmentation

Finally, contour detection and segmentation performances are studied for several levels of shape constraint. The ability of superpixel methods to provide accurate contour detection was investigated, for instance in Ref. 9, 16, 19. A superpixel decomposition can naturally produce more

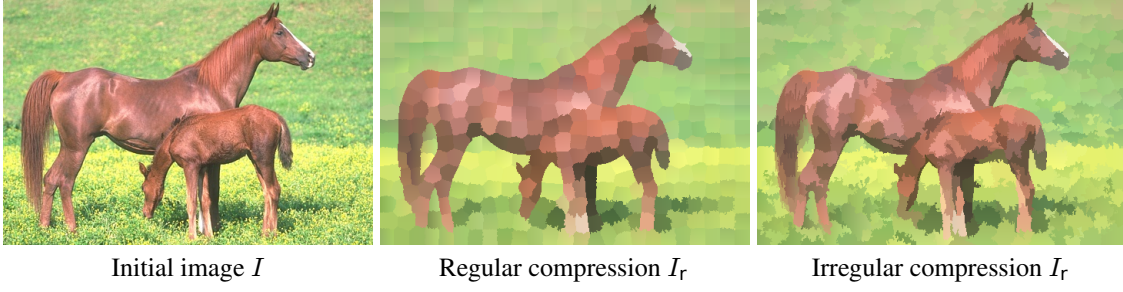


Fig 18 Example of image compression with regular and irregular decompositions for $K = 250$ superpixels. The reconstructed colors are displayed within each superpixel.

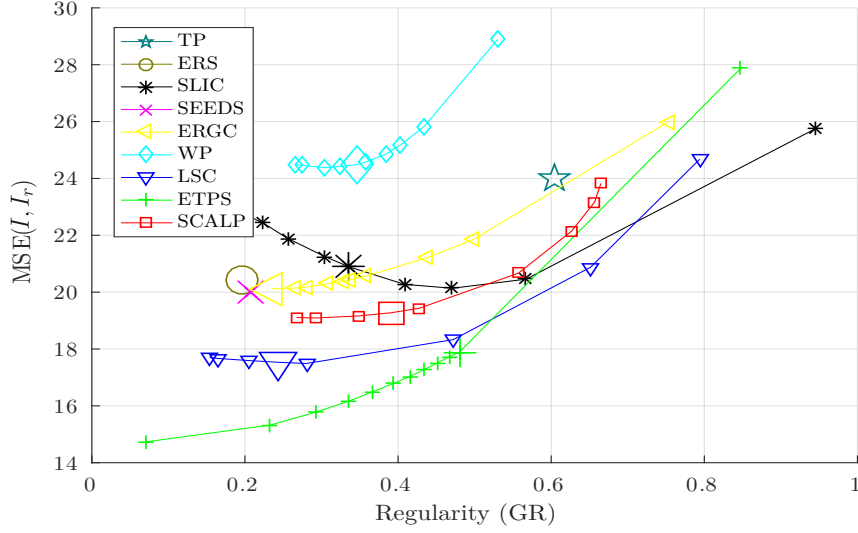


Fig 19 Results of state-of-the-art superpixel methods on image compression. The average MSE (17) between initial and reconstructed BSD images are reported.

regions than the number of image objects, and the number of superpixel boundaries is dependent on the superpixel scale. Nevertheless, accurate contour map can be easily obtained by averaging the boundaries of superpixel decompositions obtained at multiple scales.^{9,16} Hence, the average map has high values at contour pixels that are detected at several scales and is robust to texture areas and high local gradients.

As stated in Section 2.2, the boundary recall (BR) measure is not sufficient to express the contour detection performance, and should be compared to the precision (P) measure, which evaluates the percentage of accurate detection among the computed superpixel contour. Such contour

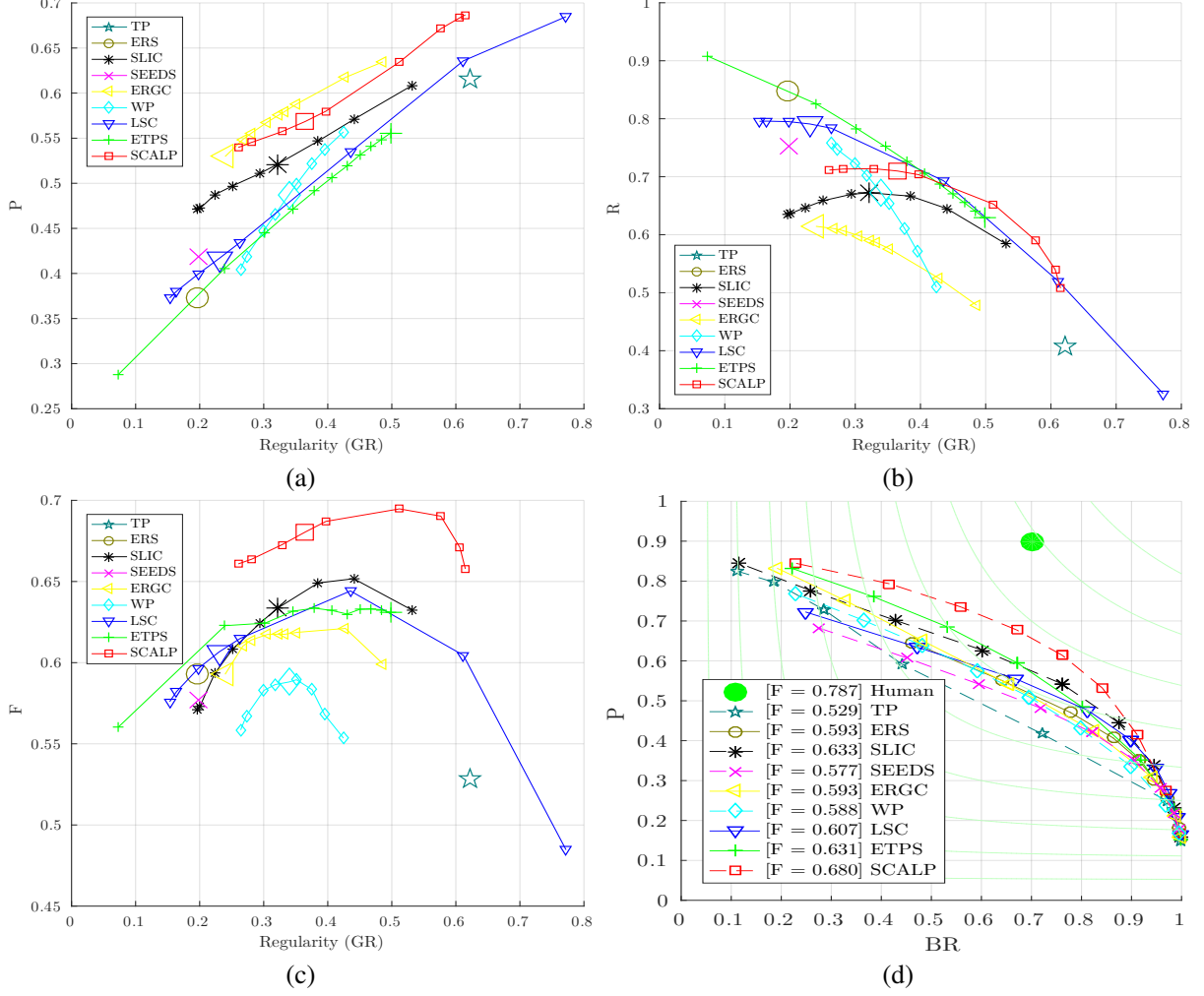


Fig 20 Contour detection performances according to the regularity with the precision (a), boundary recall (b), and maximum F-measure (c). Precision-recall curves for default regularity setting are illustrated in (d).

detection accuracy can be synthesized by the F-measure defined as:

$$F = \frac{2 \cdot P \cdot BR}{P + BR}. \quad (18)$$

In the following, superpixel boundaries are computed at several scales, from $K = [25, 1000]$, and averaged to compute the contour map. In Figure 20(d), we report the precision-boundary recall curves³⁶ corresponding to the contour detection obtained with the default regularity parameter of each method. SCALP¹⁶ and SLIC⁸ appear to obtain the best contour detection performances.

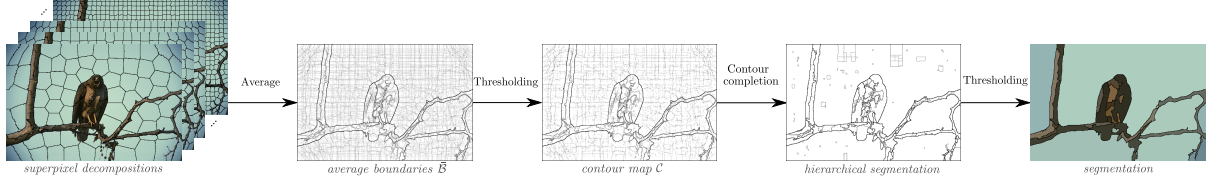


Fig 21 Segmentation from superpixel contour detection pipeline. Superpixel decompositions are computed at several scales and their boundaries are averaged to produce a contour map. The low confidence boundary pixels are removed with a thresholding. From this contour map, a hierarchical segmentation can be computed and thresholded to provide an image segmentation, that we illustrate with average colors.

To underline the impact of the regularity parameter on contour detection performances, the precision (P), boundary recall (BR), and corresponding maximum F-measure are computed for several regularity levels respectively in Figures 20(a), (b) and (c). The regularity clearly impacts the performances, since the precision and boundary recall respectively increases and drops with the GR measure in Figures 20(a) and (b). Consequently, a good trade-off for maximum F-measure is obtained for GR values between 0.35 and 0.45. More importantly, results appear to be significantly impacted by the regularity, so most methods could perform better than with their default parameters, on this application, with appropriate settings.

Finally, a segmentation can be generated using a contour completion step that compute closed regions from the contour map.⁵⁴ The whole contour detection and segmentation pipeline is illustrated in Figure 21, and we show in Figure 22 contour detection and segmentation examples computed from regular and irregular decompositions with Ref. 8. The average of regular superpixel boundaries does not enable to efficiently capture the ground truth contours (see Figure 22(c)). The contour closure of the image objects is not robust enough to lead to a segmented region with Ref. 54. With irregular decompositions, the superpixel shapes are less constrained and can better adhere to the image contours. The contour closure obtained by the average of boundaries is hence more robust and can lead to accurate segmentation (see Figure 22(b)). Nevertheless, note that the

thresholding enables to remove the boundary artifacts generated with irregular decompositions, that can obtain lower performances in terms of respect of image objects when evaluated with ASA.

5.5 Correlation between GR and Performances

In this section, we quantitatively show the correlation between the proposed global regularity measure GR and the performances of superpixel metrics and applications. We report in Table 3, the Pearson correlation coefficient between all regularity measures: C (9),³² SRC (14), J (10),¹² SMF (15) and GR (16), and the performances of ASA (6), UE (5), BR (7), P (18), EV (2) and MSE (17). The correlations are averaged on the decompositions of SLIC,⁸ ERGC,¹¹ WP,¹² LSC,¹³ ETPS¹⁴ and SCALP¹⁶ methods, which enable to set a regularity parameter, and are computed for $K = 250$ superpixels.

Logically the ASA, BR, P and EV measures are negatively correlated while the UE and MSE measures are in line with the regularity. The proposed SRC metric globally improves the correlations compared to the circularity metric C, and the proposed SMF improves the ones of the J measure. Overall, the proposed GR measure obtains the highest correlation with the performances, demonstrating its relevance compared to previously introduced metrics. Finally, note that the correlation between ASA and UE is in average -0.9969 , which further assess the assumption made in Section 2.2.

6 Discussion

Since the superpixel representation has become popular with works such as Ref. 1, many works have proposed decomposition methods using various techniques. In the addition to the high number of superpixel works, specific metrics dedicated to superpixel evaluation were progressively

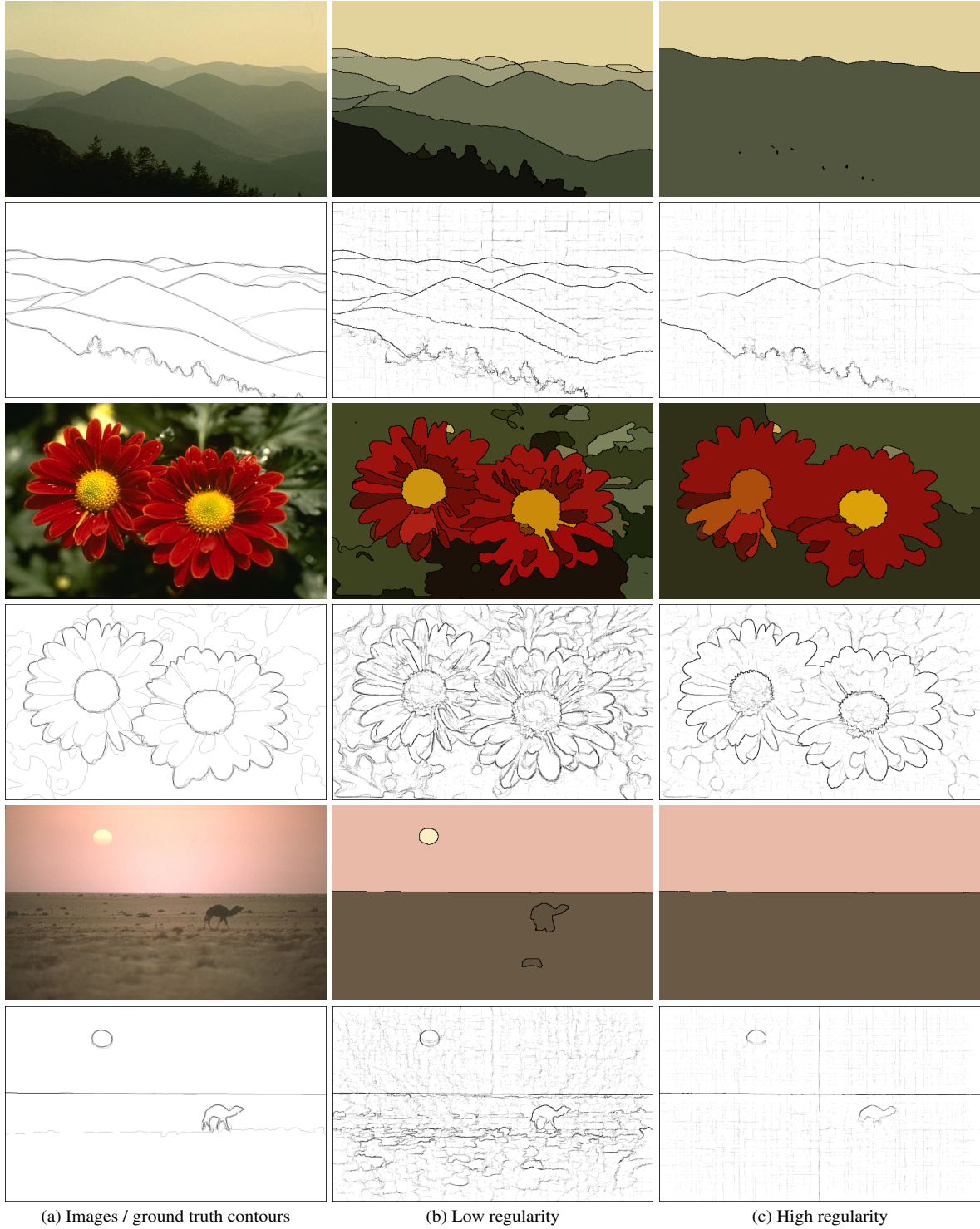


Fig 22 Examples of contour detection and segmentation with low (b) and high (c) regularity settings. Initial images with the average ground truth contours are shown in (a). Regions with average colors are illustrated after a thresholding of the hierarchical segmentation computed from the contour map with Ref. 54. The average GR is respectively 0.256 and 0.481 for the irregular and regular decompositions.

Table 3 Comparison of Pearson correlation coefficient between performances of metrics and applications, and the different regularity measures. The average results are computed on the absolute value of the correlations.

	GR (16)	SMF (15)	J (10)	SRC (14)	C (9)
ASA (6)	−0.5473	−0.5250	−0.5266	−0.5350	−0.5318
UE (5)	0.5506	0.5284	0.5299	0.5384	0.5353
BR (7)	−0.9136	−0.8974	−0.8972	−0.9049	−0.9034
P (18)	−0.9627	−0.9645	−0.9656	−0.9688	−0.9712
EV (2)	−0.6641	−0.6426	−0.6428	−0.6528	−0.6503
MSE (17)	0.6760	0.6552	0.6554	0.6655	0.6636
Average	0.8165	0.8113	0.8076	0.8122	0.8085

introduced and non exhaustively reported in the associated works. More generally, to evaluate the improvement over state-of-the-art of new proposed methods, it is crucial to have a consistent and rigorous evaluation of the reported results. Otherwise, we cannot relevantly assess the potential breakthrough performances of a new method. These issues of evaluation metrics finds an echo in other image processing applications. For instance, in most of the denoising literature, the SSIM is reported along with the PSNR metric to relevantly compare the performances, since they are complementary.

In the superpixel context, most recent methods only require to set the number of generated superpixels and to tune a regularity parameter. By reporting results evaluated according to both aspects, on the advocated metrics, we ensure that the comparison is not biased by the set parameter, and it more globally expresses the range of possible decompositions that can be obtained. Moreover, the main aim of superpixels is to be used as a pre-processing to speed up the computational time of a given application. The number of superpixels corresponds to the number of elements to process, and is generally in line with the complexity of the application. Therefore, the number of desired superpixels is set by the time constraint, and the main parameter to study is the regularity.

The evaluation of performances according to the GR measure offers a new and relevant inter-

pretation of the potential of superpixel methods, since it is directly correlated with the results of image processing applications. We quantitatively demonstrate in Section 5.5, that our measure is more correlated to the application performances than previously introduced regularity metrics. In Section 5, we show that regular decompositions may better perform for applications such as nearest neighbor matching or tracking. The adjacency relations between regular superpixels are more consistent between images, which enables to robustly describe superpixels and their neighborhood to match similar patterns.²⁴ When processing the image itself, with applications such as image compression or contour detection, more irregular decompositions may be desirable. By relaxing the regularity constraint, the superpixels group more homogeneous pixels in terms of color, which lead to more accurate compression, and more efficiently follows the image contours, leading to higher contour detection. Consequently, according to the tackled application, we demonstrate that the regularity must be controlled to provide the best possible results.

As shown throughout the paper, and illustrated in Figure 2, the same method can produce decompositions of various regularity, and perform well on several applications. Therefore, although many papers make a clear distinction between regular and irregular methods, generally according to the results produced with their default parameter, it appears that this property does not reflect the potential behavior of a method. The distinction between regular and irregular methods should be based on the method ability to produce superpixels of different sizes, instead of irregular shapes. For instance, methods such as Ref. 2, 15 can produce regions of very different sizes, while SLIC,⁸ and most methods that are based on, can produce irregular shaped superpixels by relaxing the regularity constraint, but the clustering is still constrained in a square bounding box, so the produced superpixels have consistent sizes.

7 Conclusion

In this work, we propose an evaluation framework of superpixel decomposition methods. We take a global view of existing superpixel metrics, and investigate their limitations to propose a relevant comparison process. We reduce the evaluation to the three main aspects of a decomposition: the homogeneity of color clustering with EV, the respect of the image objects with ASA, and the regularity with the proposed GR. Contrary to existing regularity metrics, the proposed GR measure relevantly evaluates both shape regularity and consistency of the superpixels, addressing the non-robustness of state-of-the-art metrics.

Since most of the recent methods enable to set a shape parameter, which impacts the decomposition performances, we also compare methods according to their regularity, evaluated with the GR measure. Hence, the reported performances are not biased by the methods settings. The GR metric is highly correlated to the performances of applications such as matching, tracking or contour detection, where the same method can perform well on several applications, according to the chosen regularity parameter. Therefore, we advocate to use the proposed framework and metrics when comparing superpixel or supervoxel methods, and to evaluate the performances at several regularity levels evaluated with GR.

Appendix A: Correlation between ASA and UE

In this section, we demonstrate the relation (8) for several assumptions.

Proposition 1. If all superpixels S_k have a major overlap, *i.e.*, if $\forall S_k, \exists G_i, |S_k \cap G_i| \geq |S_k|/2$, then, $\text{ASA}(\mathcal{S}, \mathcal{G}) = 1 - \text{UE}(\mathcal{S}, \mathcal{G})/2$.

Proof. The major overlap hypothesis, illustrated in Figure 6(b), is that each superpixel S_k has a major overlap with a region G_i that covers more than half of its area, *i.e.*,

$$\forall S_k, \exists G_i, |S_k \cap G_i| \geq |S_k|/2. \quad (19)$$

Hence, we have $\min\{|S_k \cap G_i|, |S_k \setminus G_i|\} = |S_k \setminus G_i|$, and $\forall G_{j,j \neq k}, \min\{|S_k \cap G_j|, |S_k \setminus G_j|\} = |S_k \cap G_j|$.

Therefore, we obtain,

$$\begin{aligned} \sum_{G_j} \min\{|S_k \cap G_j|, |S_k \setminus G_j|\} &= |S_k \setminus G_i| + \sum_{G_j, j \neq k} |S_k \cap G_j|, \\ &= (|S_k| - |S_k \cap G_i|) + (|S_k| - |S_k \cap G_i|), \\ &= 2(|S_k| - |S_k \cap G_i|), \\ &= 2(|S_k| - \max_{G_j} |S_k \cap G_j|). \end{aligned}$$

Consequently, we have,

$$\begin{aligned} \text{UE}(\mathcal{S}, \mathcal{G}) &= \frac{1}{|I|} \sum_{S_k} \sum_{G_j} \min\{|S_k \cap G_j|, |S_k \setminus G_j|\}, \\ &= \frac{2}{|I|} \sum_{S_k} \left(|S_k| - \max_{G_j} |S_k \cap G_j| \right), \\ &= 2 \left(1 - \frac{1}{|I|} \sum_{S_k} \max_{G_j} |S_k \cap G_j| \right), \\ &= 2(1 - \text{ASA}(\mathcal{S}, \mathcal{G})). \end{aligned}$$

Finally, according to the ASA definition (6), the relation (8) is verified. \square

Proposition 2. If all superpixels S_k overlap with at most 2 regions, *i.e.*, if $\forall S_k, |I_k| \leq 2$, with $\sum_{i \in I_k} |S_k \cap G_i| = |S_k|$, then, $\text{ASA}(\mathcal{S}, \mathcal{G}) = 1 - \text{UE}(\mathcal{S}, \mathcal{G})/2$.

Proof. In this case, a superpixel S_k can overlap with one or two regions. If S_k overlaps with only one region, *e.g.*, G_i , then the assumption in (19) is verified. Otherwise, if S_k overlaps with two regions, (19) is also necessarily true, since one of the two regions overlaps with at least half of S_k . Hence, in both cases, (8) is verified. \square

Proposition 3. If \mathcal{G} is a binary ground truth, *i.e.*, $|\mathcal{G}| \leq 2$, then, $\text{ASA}(\mathcal{S}, \mathcal{G}) = 1 - \text{UE}(\mathcal{S}, \mathcal{G})/2$.

Proof. If \mathcal{G} is a binary ground truth, it is composed of only two regions such that $\mathcal{G} = G_1 \cup G_2$, and all superpixels necessarily overlap with at most 2 regions. The assumption of Proposition 2 is hence true and (8) is verified. \square

Figure 23 demonstrates the relevance of relation (8). Very low differences are reported between this model and measured ASA and UE for decompositions of the following state-of-the-art methods: TP,⁵ ERS,⁷ SLIC,⁸ SEEDS,⁹ ERGC,¹¹ WP,¹² LSC,¹³ ETPS¹⁴ and SCALP,¹⁶ applied to the test images of the Berkeley segmentation dataset (BSD).⁴⁸ Note that the error amplitude logically reduces with the number of superpixels, *i.e.*, the accuracy of the decomposition.

Acknowledgments

This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the GOTMI project (ANR-16-CE33-0010-01) and the Investments for the future Program IdEx Bordeaux (ANR-10-IDEX-03-02) with the Cluster of excellence CPU.

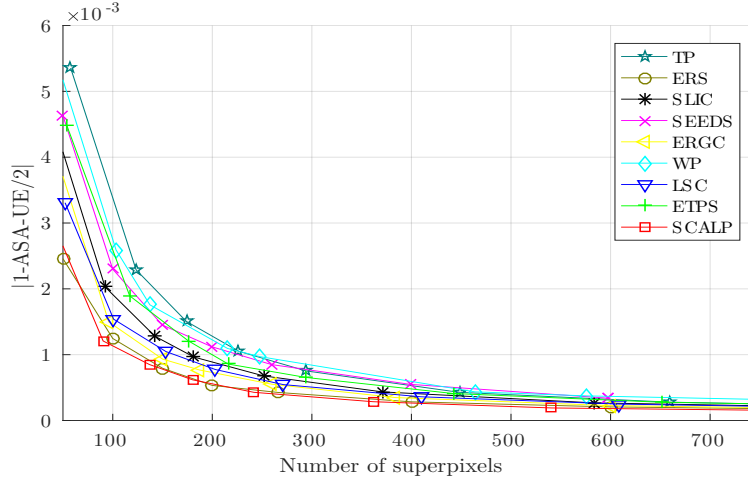


Fig 23 Error between measured ASA and UE and the relation (8). The low values compared to ASA and UE metrics demonstrate the relevance of our hypothesis.

References

- 1 X. Ren and J. Malik, “Learning a classification model for segmentation,” in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 10–17 (2003).
- 2 P. F. Felzenszwalb and D. P. Huttenlocher, “Efficient graph-based image segmentation,” *International Journal of Computer Vision (IJCV)* **59**(2), 167–181 (2004).
- 3 A. Vedaldi and S. Soatto, “Quick shift and kernel methods for mode seeking,” in *Proc. of European Conference on Computer Vision (ECCV)*, 705–718 (2008).
- 4 A. P. Moore, S. J. D. Prince, J. Warrell, *et al.*, “Superpixel lattices,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 1–8 (2008).
- 5 A. Levinstein, A. Stere, K. N. Kutulakos, *et al.*, “Turbopixels: Fast superpixels using geometric flows,” *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* **31**(12), 2290–2297 (2009).
- 6 O. Veksler, Y. Boykov, and P. Mehrani, “Superpixels and supervoxels in an energy optimiza-

- tion framework,” in *Proc. of European Conference on Computer Vision (ECCV)*, 211–224 (2010).
- 7 M. Y. Liu, O. Tuzel, S. Ramalingam, *et al.*, “Entropy rate superpixel segmentation,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2097–2104 (2011).
 - 8 R. Achanta, A. Shaji, K. Smith, *et al.*, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* **34**(11), 2274–2282 (2012).
 - 9 M. Van den Bergh, X. Boix, G. Roig, *et al.*, “SEEDS: Superpixels extracted via energy-driven sampling,” in *Proc. of European Conference on Computer Vision (ECCV)*, 13–26 (2012).
 - 10 C. Conrad, M. Mertz, and R. Mester, “Contour-relaxed superpixels,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 280–293 (2013).
 - 11 P. Buysens, M. Toutain, A. Elmoataz, *et al.*, “Eikonal-based vertices growing and iterative seeding for efficient graph-based segmentation,” in *Proc. of IEEE International Conference on Image Processing (ICIP)*, 4368–4372 (2014).
 - 12 V. Machairas, M. Faessel, D. Cárdenas-Peña, *et al.*, “Waterpixels,” *IEEE Trans. on Image Processing (TIP)* **24**(11), 3707–3716 (2015).
 - 13 Z. Li and J. Chen, “Superpixel segmentation using linear spectral clustering,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 1356–1363 (2015).
 - 14 J. Yao, M. Boben, S. Fidler, *et al.*, “Real-time coarse-to-fine topologically preserving segmentation,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2947–2955 (2015).

- 15 A. Rubio, L. Yu, E. Simo-Serra, *et al.*, “BASS: Boundary-aware superpixel segmentation,” in *Proc. of International Conference on Pattern Recognition (ICPR)*, 2824–2829 (2016).
- 16 R. Giraud, V.-T. Ta, and N. Papadakis, “Robust superpixels using color and contour features along linear path,” *HAL preprint <https://hal.archives-ouvertes.fr/hal-01510063>* (2017).
- 17 B. Fulkerson, A. Vedaldi, and S. Soatto, “Class segmentation and object localization with superpixel neighborhoods,” in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 670–677 (2009).
- 18 Y. Zhang, R. Hartley, J. Mashford, *et al.*, “Superpixels, occlusion and stereo,” in *Proc. of Int. Conf. on Digital Image Computing Techniques and Applications (DICTA)*, 84–91 (2011).
- 19 P. Arbelaez, M. Maire, C. Fowlkes, *et al.*, “Contour detection and hierarchical image segmentation,” *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* **33**(5), 898–916 (2011).
- 20 S. Gould, J. Rodgers, D. Cohen, *et al.*, “Multi-class segmentation with relative location prior,” *International Journal of Computer Vision (IJCV)* **80**(3), 300–316 (2008).
- 21 Y. Yang, S. Hallman, D. Ramanan, *et al.*, “Layered object detection for multi-class segmentation,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 3113–3120 (2010).
- 22 J. Tighe and S. Lazebnik, “SuperParsing: Scalable nonparametric image parsing with superpixels,” in *Proc. of European Conference on Computer Vision (ECCV)*, 352–365 (2010).
- 23 S. Gould, J. Zhao, X. He, *et al.*, “Superpixel graph label transfer with learned distance metric,” in *Proc. of European Conference on Computer Vision (ECCV)*, 632–647 (2014).

- 24 R. Giraud, V.-T. Ta, A. Bugeau, *et al.*, “SuperPatchMatch: An algorithm for robust correspondences using superpixel patches,” *IEEE Trans. Image Processing (TIP)* (*in press*) (2017).
- 25 R. Sawhney, F. Li, and H. I. Christensen, “GASP: Geometric association with surface patches,” in *Proc. of International Conference on 3D Vision*, 107–114 (2014).
- 26 A. Kae, K. Sohn, H. Lee, *et al.*, “Augmenting CRFs with Boltzmann machine shape priors for image labeling,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019–2026 (2013).
- 27 F. Liu, C. Shen, and G. Lin, “Deep convolutional neural fields for depth estimation from a single image,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 5162–5170 (2015).
- 28 R. Gadde, V. Jampani, M. Kiefel, *et al.*, “Superpixel convolutional networks using bilateral inceptions,” in *Proc. of European Conference on Computer Vision (ECCV)*, 597–613 (2016).
- 29 J. Chang, D. Wei, and J. W. Fisher, “A video representation using temporal superpixels,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2051–2058 (2013).
- 30 M. Reso, J. Jachalsky, B. Rosenhahn, *et al.*, “Temporally consistent superpixels,” in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 385–392 (2013).
- 31 P. Neubert and P. Protzel, “Compact watershed and preemptive SLIC: On improving trade-offs of superpixel segmentation algorithms,” in *Proc. of International Conference on Pattern Recognition (ICPR)*, 996–1001 (2014).
- 32 A. Schick, M. Fischer, and R. Stiefelhagen, “Measuring and evaluating the compactness of

- superpixels,” in *Proc. of International Conference on Pattern Recognition (ICPR)*, 930–934 (2012).
- 33 P. Neubert and P. Protzel, “Superpixel benchmark and comparison,” in *Forum Bildverarbeitung*, 1–12 (2012).
 - 34 M. Wang, X. Liu, Y. Gao, *et al.*, “Superpixel segmentation: A benchmark,” *Signal Processing: Image Communication* **56**, 28–39 (2017).
 - 35 W. Benesova and M. Kottman, “Fast superpixel segmentation using morphological processing,” in *Proc. of the Int. Conf. on Machine Vision and Machine Learning (MVML)*, (2014).
 - 36 D. R. Martin, C. C. Fowlkes, and J. Malik, “Learning to detect natural image boundaries using local brightness, color, and texture cues,” *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* **26**(5), 530–549 (2004).
 - 37 D. Stutz, A. Hermans, and B. Leibe, “Superpixels: An evaluation of the state-of-the-art,” *Computer Vision and Image Understanding (CVIU)* (*in press*) (2017).
 - 38 J. Strassburg, R. Grzeszick, L. Rothacker, *et al.*, “On the influence of superpixel methods for image parsing,” in *Proc. of the Inf. Conf. on Computer Vision Theory and Applications (VISAPP)*, 518–527 (2015).
 - 39 P. Wang, G. Zeng, R. Gan, *et al.*, “Structure-sensitive superpixels via geodesic distance,” *International Journal of Computer Vision (IJCV)* **103**(1), 1–21 (2013).
 - 40 Y. Zhang, X. Li, X. Gao, *et al.*, “A simple algorithm of superpixel segmentation with boundary constraint,” *IEEE Trans. on Circuits and Systems for Video Technology (TCSVT)* (99) (2016).

- 41 A. Schick, M. Fischer, and R. Stiefelhagen, “An evaluation of the compactness of superpixels,” *Pattern Recognition Letters* **43**, 71–80 (2014).
- 42 N. Strachan, P. Nesvadba, and A. R. Allen, “Fish species recognition by shape analysis of images,” *Pattern Recognition* **23**(5), 539–544 (1990).
- 43 P. Jaccard, “Bulletin de la société vaudoise des sciences naturelles,” *Étude comparative de la distribution florale dans une portion des Alpes et des Jura* **37**, 547–579 (1901).
- 44 R. Giraud, V.-T. Ta, and N. Papadakis, “Robust shape regularity criteria for superpixel evaluation,” *Proc. of IEEE International Conference on Image Processing (ICIP)* (2017).
- 45 V. Caselles, G. Facciolo, and E. Meinhardt, “Anisotropic cheeger sets and applications,” *Journal on Imaging Sciences (SIAM)* **2**(4), 1211–1254 (2009).
- 46 T. Roussillon, I. Sivignon, and L. Tougne, “Measure of circularity for parts of digital boundaries and its fast computation,” *Pattern Recognition* **43**(1), 37–46 (2010).
- 47 S. Tanimoto and T. Pavlidis, “A hierarchical data structure for picture processing,” *Computer graphics and image processing* **4**(2), 104–119 (1975).
- 48 D. Martin, C. Fowlkes, D. Tal, *et al.*, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, **2**, 416–423 (2001).
- 49 S. Gould, R. Fulton, and D. Koller, “Decomposing a scene into geometric and semantically consistent regions,” in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 1–8 (2009).
- 50 K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, *et al.*, “Parsing clothing in fashion photographs,”

- in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 3570–3577 (2012).
- 51 L. Vincent and P. Soille, “Watersheds in digital spaces: an efficient algorithm based on immersion simulations,” *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* **13**(6), 583–598 (1991).
- 52 S. Wang, H. Lu, F. Yang, *et al.*, “Superpixel tracking,” in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 1323–1330 (2011).
- 53 D. Tsai, M. Flagg, A. Nakazawa, *et al.*, “Motion coherent tracking using multi-label mrf optimization,” *International Journal of Computer Vision (IJCV)* **100**(2), 190–202 (2012).
- 54 P. Arbelaez, M. Maire, C. Fowlkes, *et al.*, “From contours to regions: An empirical evaluation,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2294–2301 (2009).

Rémi Giraud received the M.Sc. in telecommunications at ENSEIRB-MATMECA School of Engineers, and the M.Sc. in signal and image processing from the University of Bordeaux, France, in 2014. Since, he is pursuing his Ph.D. at Laboratoire Bordelais de Recherche en Informatique in the field of image processing. His research areas mainly include computer vision and image processing applications with non-local methods and superpixel representation, applied to natural and medical images.

Vinh-Thong Ta received the M.Sc. and Doctoral degrees in computer science from the University of Caen Basse-Normandie, France, in 2004 and 2009, respectively. From 2009 to 2010, he was an Assistant Professor in computer science with the School of Engineers of Caen, France. Since

2010, he is an Associate Professor with the Computer Science Department, School of Engineers ENSEIRB-MATMECA. His research mainly concerns image and data processing.

Nicolas Papadakis received the Graduate degree in applied mathematics from the National Institute of Applied Sciences, Rouen, France, in 2004, and the Ph.D. degree in applied mathematics from the University of Rennes, France, in 2007. He is currently a Researcher from the Centre National de la Recherche Scientifique, Institut de Mathématiques de Bordeaux, France. His main research interests include tracking, motion estimation, and optimal transportation for image processing and data assimilation problems.

List of Figures

- 1 Examples of trade-off between the superpixel properties for the decomposition of images (a) and (d) into three superpixels. The decomposition in (b) and (e), are optimal in terms of color homogeneity while the ones in (c) and (f) are respectively optimal in terms of respect of image contours and regularity.
- 2 Example of decompositions obtained with the same method⁸ with different regularity settings. The difference between the decompositions can be expressed with the variance of distances between superpixel barycenters (normalized by the average distance), which is 0.632 for the irregular decomposition and 0.406 for the regular one.
- 3 Comparison of ICV and EV to measure the superpixel color homogeneity. The dynamic and dimension of the same image (a) are respectively modified in (b) and (c). EV is robust to these transformations contrary to ICV.

- 4 Example of superpixel decomposition (a), with average colors (b), and color variance $\sigma(S_k)^2$ within each superpixel S_k (c).
- 5 Examples of irregular¹⁴ (b) and regular⁸ (c) decomposition of an image (a), with maximal BR measure. BR evaluates the detection of ground truth contours so both decompositions can have maximal BR measure, although the irregular one produces more superpixel boundaries.
- 6 Examples of decomposition where a superpixel S_k overlap with multiple ground truth regions G_j . In (b), contrary to (a), a region (G_2) overlaps with more than half of S_k , which corresponds to the major overlap hypothesis ensuring (8).
- 7 Illustration of the average superpixel shape definition.¹² A decomposition is considered in (a). The superpixel shapes are registered and averaged into S^* (b) to provide the binary average shape (c).
- 8 Comparison of circularity (C) and proposed shape regularity criteria (SRC) on three groups of synthetic shapes with smooth (top) and noisy boundaries (bottom). C appears to only favor circular shapes and does not separate standard and regular noisy ones. The SRC metric addresses these issues and clearly differentiates the shape groups in the smooth and noisy cases.
- 9 Robustness to the superpixel scale of the proposed shape regularity criteria (SRC) compared to the circularity (C).
- 10 Limitation of the local shape regularity metrics. The shape consistency is not considered so the quadtree decomposition gives higher measures than the decomposition obtained with the SLIC method.⁸ The decompositions are represented with their Delaunay graphs, connecting the barycenters of adjacent superpixels.

- 11 Illustration of several decomposition examples with the corresponding average binary shape in red, and J (10) and SMF (15) values.
- 12 Decomposition examples with corresponding J (10) and SMF (15) measures.
- 13 Decomposition example of each considered state-of-the-art superpixel methods for approximately 250 superpixels.
- 14 Evaluation of state-of-the-art superpixel methods on EV, ASA and GR according to the number of superpixels, with the methods default regularity settings. The hierarchy between the performances of methods tends to be consistent on all metrics.
- 15 Evaluation of superpixel state-of-the-art superpixel methods on EV, ASA according to the regularity (GR) for $K = 250$.
- 16 Examples of matching between regular (a), (b) and irregular (d), (e) decompositions. The magnitude of the displacement between the barycenters of the matched superpixels is respectively given for regular and irregular matching in (c) and (f). The displacement is illustrated with the standard optical flow representation (g), and respectively displayed within the superpixel boundaries of images (a) and (d). On these examples, the average GR is respectively 0.478 and 0.205 for the regular (a) and (b) and irregular (d) and (e) decompositions.
- 17 Examples on superpixel tracking performances of TSP²⁹ over the sequences of Ref. 53 for regular and irregular decompositions. On these examples, the average GR is respectively 0.308 and 0.554 for the irregular and regular decompositions.
- 18 Example of image compression with regular and irregular decompositions for $K = 250$ superpixels. The reconstructed colors are displayed within each superpixel.

- 19 Results of state-of-the-art superpixel methods on image compression. The average MSE (17) between initial and reconstructed BSD images are reported.
- 20 Contour detection performances according to the regularity with the precision (a), boundary recall (b), and maximum F-measure (c). Precision-recall curves for default regularity setting are illustrated in (d).
- 21 Segmentation from superpixel contour detection pipeline. Superpixel decompositions are computed at several scales and their boundaries are averaged to produce a contour map. The low confidence boundary pixels are removed with a thresholding. From this contour map, a hierarchical segmentation can be computed and thresholded to provide an image segmentation, that we illustrate with average colors.
- 22 Examples of contour detection and segmentation with low (b) and high (c) regularity settings. Initial images with the average ground truth contours are shown in (a). Regions with average colors are illustrated after a thresholding of the hierarchical segmentation computed from the contour map with Ref. 54. The average GR is respectively 0.256 and 0.481 for the irregular and regular decompositions.
- 23 Error between measured ASA and UE and the relation (8). The low values compared to ASA and UE metrics demonstrate the relevance of our hypothesis.

List of Tables

- 1 Average EV, ASA and GR on several scales $K = [25, 1000]$, with the methods default regularity settings.

- 2 Tracking accuracy with TSP²⁹ on sequences from Ref. 53. The labeling accuracy and the percentage of lost superpixels over frames are reported for regular and irregular settings.
- 3 Comparison of Pearson correlation coefficient between performances of metrics and applications, and the different regularity measures. The average results are computed on the absolute value of the correlations.