Robust Superpixels using Color and Contour Features along Linear Path

Rémi Giraud, Vinh-Thong Ta, Nicolas Papadakis

▶ **To cite this version:**

Rémi Giraud, Vinh-Thong Ta, Nicolas Papadakis. Robust Superpixels using Color and Contour Features along Linear Path. Computer Vision and Image Understanding, 2018. hal-01510063

HAL Id: hal-01510063
https://hal.science/hal-01510063

Submitted on 19 Jan 2018

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.

Sujet : CVIU-17-392R2: Final Decision
De : "CVIU (ELS)" <eesserver@eesmail.elsevier.com>
Date : 15/01/2018 14:45
Pour : rgiraud@labri.fr
Copie à : olga@csd.uwo.ca, nikos.paragios@ecp.fr

Ms. No.: CVIU-17-392R2
Title: Robust superpixels using color and contour features along linear path
Corresponding Author: Mr. Rémi Giraud
Authors: Vinh-Thong Ta; Nicolas Papadakis

Dear Mr. Giraud,

I am pleased to inform you that your manuscript referenced above has been accepted for publication in Computer Vision and Image Understanding.
Your accepted manuscript will now be transferred to our production department and work will begin on creation of the proof. If we need any additional information to create the proof, we will let you know. If not, you will be contacted again in the next few days with a request to approve the proof and to complete a number of online forms that are required for publication.

When your paper is published on ScienceDirect, you want to make sure it gets the attention it deserves. To help you get your message across, Elsevier has developed a new, free service called AudioSlides: brief, webcast-style presentations that are shown (publicly available) next to your published article. This format gives you the opportunity to explain your research in your own words and attract interest. You will receive an invitation email to create an AudioSlides presentation shortly. For more information and examples, please visit http://www.elsevier.com/audioslides.

Thank you for your contribution to Computer Vision and Image Understanding.

With kind regards,

Nikos Paragios
Editor-in-Chief
Computer Vision and Image Understanding

Editorial/Production Office
Elsevier
525 B Street, Suite 1800
San Diego, CA 92101-4495
USA
Phone: (619) 699-6484
E-mail: cviu@elsevier.com

Computer Vision and Image Understanding
journal homepage: www.elsevier.com

Robust superpixels using color and contour features along linear path

Rémi Giraud[a,b,c,d,**], Vinh-Thong Ta[a,b,e], Nicolas Papadakis[c,d]

[a] *Univ. Bordeaux, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.*
[b] *CNRS, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.*
[c] *Univ. Bordeaux, IMB, UMR 5251, F-33400 Talence, France.*
[d] *CNRS, IMB, UMR 5251, F-33400 Talence, France.*
[e] *Bordeaux INP, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.*

ABSTRACT

Superpixel decomposition methods are widely used in computer vision and image processing applications. By grouping homogeneous pixels, the accuracy can be increased and the decrease of the number of elements to process can drastically reduce the computational burden. For most superpixel methods, a trade-off is computed between 1) color homogeneity, 2) adherence to the image contours and 3) shape regularity of the decomposition. In this paper, we propose a framework that jointly enforces all these aspects and provides accurate and regular Superpixels with Contour Adherence using Linear Path (SCALP). During the decomposition, we propose to consider color features along the linear path between the pixel and the corresponding superpixel barycenter. A contour prior is also used to prevent the crossing of image boundaries when associating a pixel to a superpixel. Finally, in order to improve the decomposition accuracy and the robustness to noise, we propose to integrate the pixel neighborhood information, while preserving the same computational complexity. SCALP is extensively evaluated on standard segmentation dataset, and the obtained results outperform the ones of the state-of-the-art methods. SCALP is also extended for supervoxel decomposition on MRI images.
© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

The use of superpixels has become a very popular technique for many computer vision and image processing applications such as: object localization (Fulkerson et al., 2009), contour detection (Arbelaez et al., 2011), face labeling (Kae et al., 2013), data associations across views (Sawhney et al., 2014), or multi-class object segmentation (Giraud et al., 2017b; Gould et al., 2008, 2014; Tighe and Lazebnik, 2010; Yang et al., 2010). Superpixel decomposition methods group pixels into homogeneous regions, providing a low-level representation that tries to respect the image contours. For image segmentation, where the goal is to split the image into similar regions according to object, color or texture priors, the decomposition into superpixels may improve the segmentation accuracy and decrease the computational burden. (Gould et al., 2014). Contrary to multi-resolution approaches, that decrease the image size, superpixels preserve the image geometry, since their boundaries follow the image contours. Hence, the results obtained at the superpixel level may be closer to the ground truth result at the pixel level.

Many superpixel methods have been proposed using various techniques. Although the definition of an optimal decomposition depends on the tackled application, most methods tend to achieve the following properties. First, the boundaries of the decomposition should adhere to the image contours, and superpixels should not overlap with multiple objects. Second, the

**Corresponding author: Tel.: +33(0)540006937; fax: +33(0)50006669;
e-mail: remi.giraud@labri.fr (Rémi Giraud)

superpixel clustering must group pixels with homogeneous colors. Third, the superpixels should have compact shapes and consistent sizes. The shape regularity helps to visually analyze the image decomposition and has been proven to impact application performances (Reso et al., 2013; Veksler et al., 2010; Strassburg et al., 2015). Finally, since superpixels are usually used as a pre-processing step, the decomposition should be obtained in limited computational time and allow the control of the number of produced elements.

To achieve the aforementioned properties, most state-of-the-art methods compute a trade-off between color homogeneity and shape regularity of the superpixels. Nevertheless, some approaches less consider the regularity property and can produce superpixels of highly irregular shapes and sizes. In the following, we present an overview of the most popular superpixel methods, defined as either irregular or regular ones. Note that although some methods can include terms into their models to generate for instance more regular results, *e.g.*, Van den Bergh et al. (2012), we here consider methods in their default settings, as described by the authors.

The regularity criteria can be seen as the behavior to frequently produce irregular regions, in terms of both shapes and sizes (Giraud et al., 2017c). Methods such as Felzenszwalb and Huttenlocher (2004); Vedaldi and Soatto (2008) generate very irregular regions in terms of both size and shape while SLIC can generate a few irregular shapes but their sizes are constrained into a fixed size window.

Irregular Superpixel Methods

With irregular methods, superpixels can have very different sizes and stretched shapes. For instance, small superpixels can be produced, without enough pixels to compute a significant descriptor. Too large superpixels may also overlap with several objects contained in the image. First segmentation methods, such as the watershed approach, *e.g.*, Vincent and Soille (1991), compute decompositions of highly irregular size and shape. Methods such as Mean shift (Comaniciu and Meer, 2002) or Quick shift (Vedaldi and Soatto, 2008) consider an initial decomposition and perform a histogram-based segmentation. However, they are very sensitive to parameters and are obtained with high computational cost (Vedaldi and Soatto, 2008). Another approach considers pixels as nodes of a graph to perform a faster agglomerative clustering (Felzenszwalb and Huttenlocher, 2004). These methods present an important drawback: they do not allow to directly control the number of superpixels. This is particularly an issue when superpixels are used as a low-level representation to reduce the computational time.

The SEEDS method (Van den Bergh et al., 2012) proposes a coarse-to-fine approach starting from a regular grid. However, this method may provide superpixels with irregular shapes. Although a compactness constraint can be set to compute regular superpixels, the authors report degraded results of decomposition accuracy with such approach.

Regular Superpixel Methods

For superpixel-based object recognition methods, *e.g.*, Gould et al. (2008, 2014), or video tracking, *e.g.*, Reso et al. (2013); Wang et al. (2011), the use of regular decompositions is mandatory, *i.e.*, decompositions with superpixels having approximately the same size and compact shapes. For instance, for superpixel-based video tracking applications, the tracking of object trajectories within a scene is improved with consistent decompositions over time (Chang et al., 2013; Reso et al., 2013).

Most of the regular methods consider an initial regular grid, allowing to set the number of superpixels, and update superpixels boundaries while applying spatial constraints. Classical methods are based on region growing, such as Turbopixels (Levinshtein et al., 2009) using geometric flows, or eikonal-based methods, *e.g.*, ERGC (Buyssens et al., 2014), while other approaches use graph-based energy models (Liu et al., 2011; Veksler et al., 2010). In Machairas et al. (2015), a watershed algorithm is adapted to produce regular decompositions using a spatially regularized image gradient. Similarly to SEEDS (Van den Bergh et al., 2012), a coarse-to-fine approach has recently been proposed in Yao et al. (2015), producing highly regular superpixels.

The SLIC method (Achanta et al., 2012) performs an iterative

accurate clustering, while providing regular superpixels, in order of magnitude faster than graph-based approaches (Liu et al., 2011; Veksler et al., 2010). The SLIC method has been extended in several recent works, *e.g.*, Chen et al. (2017); Huang et al. (2016); Rubio et al. (2016); Zhang et al. (2016); Zhang and Zhang (2017). However, it can fail to adhere to image contours, as for other regular methods, *e.g.*, Levinshtein et al. (2009); Yao et al. (2015), since it is based on simple local color features and globally enforces the decomposition regularity using a fixed trade-off between color and spatial distances.

Contour Constraint

In the literature, several works have attempted to improve the decomposition performances in terms of contour adherence by using gradient or contour prior information. In Mori et al. (2004), a contour detection algorithm is used to compute a pre-segmentation using the normalized cuts algorithm (Shi and Malik, 2000). The segmentation may accurately guide the superpixel decomposition, but such approaches based on normalized cuts are computationally expensive (Mori et al., 2004). Moreover, the contour adherence of the produced decompositions are far from state-of-the-art results (Achanta et al., 2012). In Moore et al. (2008), the superpixel decomposition is constrained to fit to a grid, also called superpixel lattice. The decomposition is then refined using graph cuts. However, this method is very dependent on the used contour prior. Moreover, although the superpixels have approximately the same sizes, they have quite irregular shapes and may appear visually unsatisfactory.

In Machairas et al. (2015), the image gradient information is used to constrain the superpixel boundaries, but the results on superpixel evaluation metrics are lower than the ones of SLIC (Achanta et al., 2012). In Zhang et al. (2016), the local gradient information is considered to improve the superpixel boundaries evolution. However, the computational cost of the method is increased by a 10× order of magnitude compared to SLIC.

Segmentation from Contour Detection

Contour detection methods generally do not enforce the contour closure. To produce an image segmentation, a contour completion step is hence necessary. Many contour completion methods have been proposed (see for instance Arbelaez et al. (2011) and references therein). This step may improve the accuracy of the contour detection, since objects are generally segmented by closed curves.

Methods such as Arbelaez and Cohen (2008); Arbelaez et al. (2009), propose a hierarchical image segmentation based on contour detection. This can be considered as a probability contour map, that produces a set of closed curves for any threshold. Although such methods enable to segment an image from a contour map, they do not allow to control the size, the shape and the number of the produced regions, while most superpixel methods enable to set the number of superpixels which approximately have the same size. Moreover, the performances of the contour detection is extremely dependent on the fixed threshold parameter, which depends on the image content (Arbelaez et al., 2009). Hence, they are mainly considered as segmentation methods and cannot be considered as relevant frameworks to compute superpixel decompositions.

Robustness to Noise

Superpixel decompositions are usually used as a pre-processing step in many computer vision applications. Therefore, they tend to be applied to heterogeneous images that can suffer from noise. Moreover, image textures and high local gradients may also mislead the superpixel decomposition. Most of the state-of-the-art superpixel methods are not robust to noise, and provide degraded decompositions when applied to slightly noised images or images with low resolution. With such approaches, a denoising step is necessary to compute a relevant decomposition. For instance, the watershed approach of Machairas et al. (2015) uses a pre-filtering step to smooth local gradients according to the given size of superpixels. Nevertheless, this step is only designed to smooth local gradients of initial images and the impact of this filtering is not reported (Machairas et al., 2015).

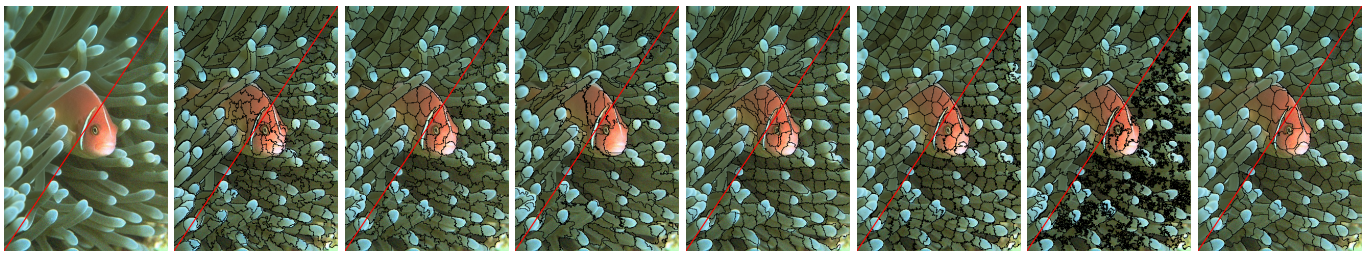


Fig. 1. Comparison of the proposed SCALP approach to the following state-of-the-art superpixel methods: ERS (Liu et al., 2011), SLIC (Achanta et al., 2012), SEEDS (Van den Bergh et al., 2012), ERGC (Buyssens et al., 2014), ETPS (Yao et al., 2015) and LSC (Chen et al., 2017). SCALP obtains the most visually satisfying result with superpixels that adhere well to the image contours. A Gaussian noise has been added to the bottom-right part of the image to demonstrate that SCALP is robust to noise, contrary to most of the compared methods.

Contributions

In this paper, we propose a method that produces accurate, regular and robust Superpixels with Contour Adherence using Linear Path (SCALP)[1]. Our decomposition approach aims to jointly improve all superpixel properties: color homogeneity, respect of image objects and shape regularity. In Figure 1, we compare the proposed approach to state-of-the-art methods on an example result. SCALP provides a more satisfying result that respects the image contours. Moreover, contrary to most state-of-the-art methods, SCALP is robust to noise, since it provides accurate and regular decompositions on the noisy part of the image.

- Most state-of-the-art methods have very degraded performances when applied to even slightly noised images (see Figure 1). We propose to consider the neighboring pixels information during the decomposition process. We show that these features can be integrated at the same computational complexity, while they improve the decomposition accuracy and the robustness to noise.

- To further enforce the color homogeneity within a regular shape, we define the linear path between the pixel and the superpixel barycenter, and we consider color features along the path. Contrary to geodesic distances that can allow irregular paths leading to non convex shapes, our linear path naturally enforces the decomposition regularity. A contour prior can also be used to enforce the respect of image objects and prevent the crossing of image contours when associating a pixel to a superpixel.

- We propose a framework to generate superpixels within an initial segmentation computed from a contour prior completion. The produced superpixels are regular in terms of size and shape although they are constrained by the segmentation to obtain higher contour adherence performances.

- We provide an extensive evaluation of SCALP on the Berkeley segmentation dataset (BSD). Our results outperform recent state-of-the-art methods, on initial and noisy images, in terms of superpixel and contour detection metrics.

- Finally, we naturally extend SCALP to supervoxel decomposition and provide results on magnetic resonance imaging (MRI) segmentation.

This paper is an extension of the work proposed in Giraud et al. (2016), with substantial new improvements such as the use in constant time of the neighboring pixels information, the use of contour prior by considering the maximum intensity on the linear path, or the extension to supervoxels. We show that these new contributions improve the decomposition performances, and by performing the clustering in a high dimensional feature space (Chen et al., 2017), SCALP substantially outperforms Giraud et al. (2016) and the recent state-of-the-art methods.

[1] An implementation of the proposed SCALP method is available at: `www.labri.fr/~rgiraud/research/scalp.php`

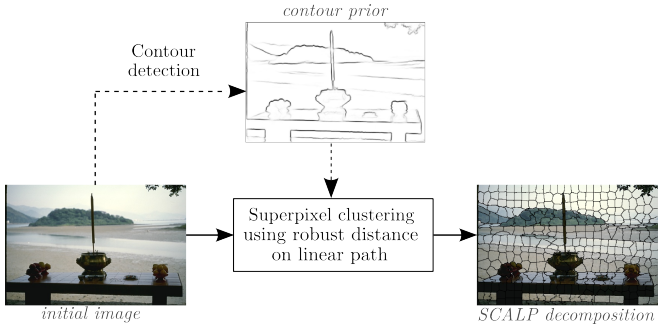


Fig. 2. The SCALP framework. A prior can be used (dotted arrows) to enforce the respect of image contours, leading to an accurate decomposition. When trying to associate a pixel to a superpixel, SCALP considers the color information from neighboring pixels, and color and contour features on the linear path to the superpixel barycenter.

2. SCALP Framework

The SCALP framework is based on the simple linear iterative clustering framework (SLIC) (Achanta et al., 2012), and is summarized in Figure 2. In this section, we first present SLIC and then propose several improvements: a robust distance on pixel neighborhood, the use of features along the linear path to the superpixel barycenter and a framework considering an initial segmentation as constraint while producing regular superpixels.

2.1. Iterative Clustering Framework

The iterative clustering framework introduced in Achanta et al. (2012) proposes a fast iterative framework using simple color features (average in CIELab colorspace). The decomposition is initialized by a regular grid with blocks of size $r\times r$. This size is computed by the ratio between the number of pixels N and the number of desired superpixels K, such that $r = \sqrt{N/K}$. A color clustering is then iteratively performed into fixed windows of size $(2r+1)\times(2r+1)$ pixels centered on the superpixel barycenter. The superpixel is thus constrained into this window, which limits its size. Each superpixel S_k is described by a cluster C_k, that contains the average CIELab color feature on pixels $p \in S_k$, $F_k = [l_k, a_k, b_k]$, and $X_k = [x_k, y_k]$, the spatial barycenter of S_k such that $C_k = [F_k, X_k]$. The iterative clustering consists, for each cluster C_k, in testing all pixels $p = [F_p, X_p]$ within a $(2r+1)\times(2r+1)$ pixels window centered on X_k, by computing a spatial distance d_s, and a color distance d_c:

$$d_s(p, C_k) = (x_p - x_k)^2 + (y_p - y_k)^2, \tag{1}$$

$$d_c(p, C_k) = (l_p - l_k)^2 + (a_p - a_k)^2 + (b_p - b_k)^2, \tag{2}$$

$$D(p, C_k) = d_c(p, C_k) + d_s(p, C_k)\frac{m^2}{r^2}, \tag{3}$$

with m the regularity parameter that sets the trade-off between spatial and color distances. High values of m produce more regular superpixels, while small values allow better adherence to image boundaries, producing superpixels of more variable sizes and shapes. The pixel p is associated to the superpixel S_k minimizing (3).

Nevertheless, since a parameter m is set to enforce the regularity in (3), SLIC can fail to both produce regular superpixels and to adhere to the image contours. In the following, we show how the decomposition accuracy can be improved with a more robust distance, by considering neighboring color features and information of pixels along the linear path to the superpixel barycenter.

2.2. Robust Distance on Pixel Neighborhood

Natural images may present high local image gradients or noise, that can highly degrade the decomposition into superpixels. In this section, we propose to consider the pixel neighborhood to improve both accuracy and robustness, and we give a method to integrate this information in the decomposition process at a constant complexity.

2.2.1. Distance on Neighborhood

We propose to integrate the neighboring pixels information in our framework when computing the clustering distance between a pixel p and a cluster C_k. Similarly to patch-based approaches, the pixels in a square area $\mathcal{P}(p)$ centered on p, of size $|\mathcal{P}(p)| = (2n+1)\times(2n+1)$ pixels, are considered in the proposed color distance D_c:

$$D_c(p, C_k) = \sum_{q\in\mathcal{P}(p)} (F_q - F_{C_k})^2 w_{p,q}. \tag{4}$$

To be robust to high local gradients while preserving the image contours, we define $w_{p,q}$ such that $w_{p,q}$ =

$\exp\left(-(F_p - F_q)^2/(2\sigma^2)\right)/Z$, with Z the normalization factor such that $Z = \sum_{q\in\mathcal{P}(p)} \exp\left(-(F_p - F_q)^2/(2\sigma^2)\right)$, and $\sum_{q\in\mathcal{P}(p)} w_{p,q} = 1$.

2.2.2. Fast Distance Computation

The complexity of the proposed distance (4) is $O(N)$, with $N = (2n+1)^2 = |\mathcal{P}(p)|$, the number of pixels in the neighborhood. We propose a method that drastically reduces the computational burden of (4). Since the distance is computed between a set of pixels and a cluster, it can be decomposed and partially pre-computed.

Proposition 1. *Eq.* (4) *can be computed at complexity* $O(1)$.

Proof. The distance between features F in (4) reads:

$$\begin{aligned}
&\sum_{q\in\mathcal{P}(p)} (F_q - F_{C_k})^2 w_{p,q} \\
&\quad= \sum_{q\in\mathcal{P}(p)} \left(F_q^2 + F_{C_k}^2 - 2F_qF_{C_k}\right)w_{p,q}, \\
&\quad= \sum_{q\in\mathcal{P}(p)} F_q^2 w_{p,q} + \sum_{q\in\mathcal{P}(p)} F_{C_k}^2 w_{p,q} - 2\sum_{q\in\mathcal{P}(p)} F_qF_{C_k} w_{p,q}, \\
&\quad= \mathcal{F}_p^{(2)} + F_{C_k}^2 \sum_{q\in\mathcal{P}(p)} w_{p,q} - 2F_{C_k} \sum_{q\in\mathcal{P}(p)} F_q w_{p,q}, \\
&\quad= \mathcal{F}_p^{(2)} + F_{C_k}^2 - 2F_{C_k}\mathcal{F}_p^{(1)}. \qquad (5)
\end{aligned}$$

In Eq. (5), the terms $\mathcal{F}_p^{(2)} = \sum_{q\in\mathcal{P}(p)} F_q^2 w_{p,q}$, and $\mathcal{F}_p^{(1)} = \sum_{q\in\mathcal{P}(p)} F_q w_{p,q}$, which only depend on the initial image, can be pre-computed at the beginning of the algorithm. The complexity of the proposed distance D_c is hence reduced to $O(1)$ instead of $O(N)$. □

2.3. Color and Contour Features on Linear Path

A superpixel decomposition is considered as satisfying according to the homogeneity of the color clustering and the respect of image contours. To enforce these aspects, we propose to consider color and contour features on the linear path between the pixel and the superpixel barycenter. We define the linear path $\mathbf{P}_p^k$, that contains the pixels starting from X_p, the position of a pixel p, to X_k, the barycenter of a superpixel S_k.

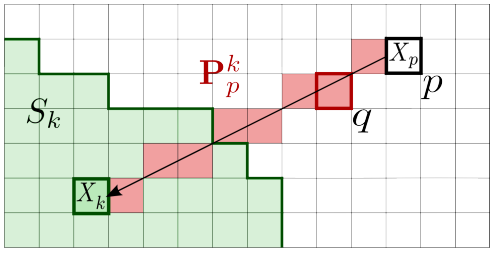


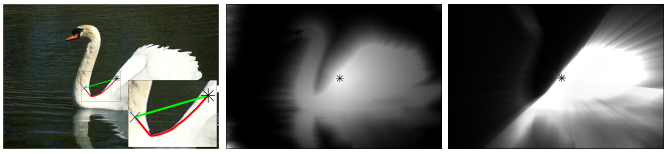
Fig. 3. Illustration of the linear path $\mathbf{P}_p^k$ between a pixel p and a superpixel S_k of barycenter X_k.

2.3.1. Linear Path between Pixel and Superpixel Barycenter

The considered linear path $\mathbf{P}_p^k$ between a pixel p and the barycenter of a superpixel S_k is illustrated in Figure 3. The pixels $q \in \mathbf{P}_p^k$ (red) are those that intersect with the segment (arrow) between X_p, the position of pixel p (black), and X_k, the barycenter of the superpixel S_k (green). Pixels q are selected such that each one only has 2 neighbors belonging to the path within a 3×3 pixels neighborhood.

Other works consider a geodesic distance to enforce the color homogeneity (Rubio et al., 2016; Wang et al., 2013) or the respect of object contours (Zhang and Zhang, 2017). The colors along the geodesic distance must be close to the average superpixel color to enable the association of the pixel to the superpixel, leading to potential irregular shapes. We illustrate this aspect in Figure 4. We compare a geodesic distance and average color distance on the linear path. While the geodesic can find a sinuous path to connect distant pixels, our linear path penalizes the crossing of regions with different colors.

A decomposition example for SCALP and a method based on a geodesic color distance (Rubio et al., 2016) is given in Figure 5. By considering the proposed linear path, we limit the computational cost, that can be substantial for geodesic distances, and we enforce the decomposition compactness, since features are considered on the direct path to the superpixel barycenter. More precisely, our linear path encourages the star-convexity property (Gulshan et al., 2010), *i.e.*, for a given shape, it exists a specific point, in our case, the superpixel barycenter, from which each point of the shape can be reached by a linear path that does not escape from the shape.



(a) Image (b) Geodesic distance (c) Linear path distance

Fig. 4. Comparison of geodesic path (red) and linear path (green) between an initial (star) and final (cross) pixel position in (a). In (b) and (c), lighter colors indicate a lower distance from the initial point (star). When trying to associate this point to a superpixel, the distance at the superpixel barycenter position is considered. Contrary to the linear path, defined in the spatial space, the geodesic path, defined in the color space, may lead to irregular superpixel shapes.

(a) Image (b) Rubio et al. (2016) (geodesic) (c) SCALP (linear path)

Fig. 5. Comparison on an image (a) of decomposition approaches using a geodesic color distance (Rubio et al., 2016) (b), and the proposed method SCALP, computing a color distance on the linear path to the superpixel barycenter (c). SCALP generates regular shapes while the geodesic-based method can create irregular superpixels.

Finally, note that despite the large number of pixel information considered during the decomposition process, the computational cost can be very limited. In practice, at a given iteration, for a given superpixel, the distance between a pixel and the superpixel has only to be computed once. The color distance can indeed be stored for each pixel and directly used for another linear path containing this pixel. Moreover, a very slight approximation can be made by directly storing for each pixel the average distance on the linear path to the superpixel barycenter, and using it when crossing an already processed pixel on a new linear path.

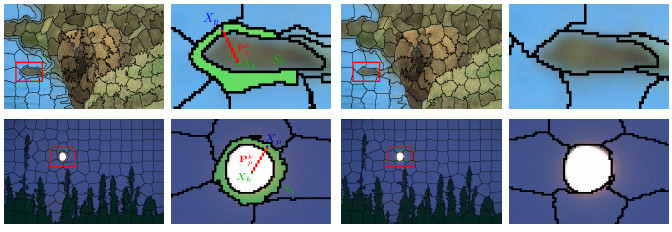
2.3.2. *Color Distance to Cluster*

The distance to minimize during the decomposition is composed of a color and a spatial term. Nevertheless, the color distance is now also computed on $\mathbf{P}_p^k$, *i.e*, between the cluster and the pixels on the linear path to the superpixel barycenter. We define the new color distance as:

$$d_c(p, C_k, \mathbf{P}_p^k) = \lambda D_c(p, C_k) + (1-\lambda)\frac{1}{|\mathbf{P}_p^k|}\sum_{q\in\mathbf{P}_p^k} D_c(q, C_k), \qquad (6)$$

where $\lambda \in [0,1]$ weights the influence of the color distance along the path. With the proposed distance (6), colors on the path to the barycenter should be close to the superpixel average color.

The distance (6) naturally enforces the regularity and also prevents irregular shapes to appear. Figure 6 shows two examples of irregular shapes that can be computed with SLIC (Achanta et al., 2012), for instance in areas of color gradation. The barycenters X_k of these irregular superpixels S_k are not contained within the shapes. The linear path $\mathbf{P}_p^k$ hence capture pixels with colors that are far from the average one of S_k. Therefore, (6) penalizes the clustering of all pixels $p \in S_k$ to this superpixel during the current iteration, so they are associated to neighboring superpixels.

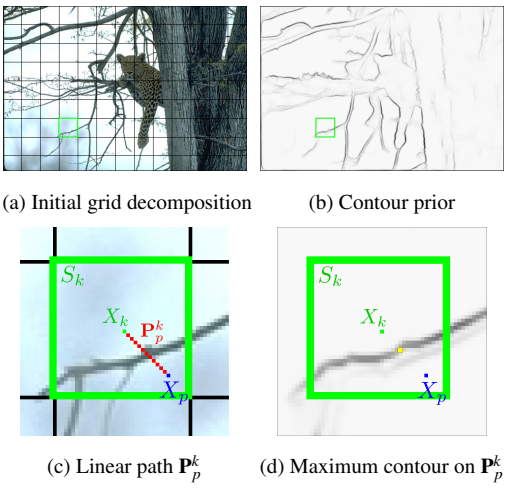


(a) SLIC irregular shapes (b) SCALP regular shapes

Fig. 6. Examples of irregular shapes obtained with SLIC (Achanta et al., 2012) (a) and regular shapes obtained with SCALP using the color distance on the linear path (6) **(b). With non regular shapes, the barycenter may fall outside the superpixel, and the linear path cross regions with different colors, penalizing the clustering distance.**

2.3.3. *Adherence to Contour Prior*

Since the optimal color homogeneity may be not in line with the respect of image objects, or fail to catch thin edges, we propose to consider the information of a contour prior map C on the linear path. Such map sets $C(p)$ to 1 if a contour is detected at pixel p, and to 0 otherwise. We propose a fast and efficient way to integrate a contour prior by weighting the distance between a pixel and a superpixel cluster by $d_C(\mathbf{P}_p^k)$, considering



(a) Initial grid decomposition (b) Contour prior

(c) Linear path $\mathbf{P}_p^k$ (d) Maximum contour on $\mathbf{P}_p^k$

Fig. 7. Illustration of SCALP first iteration starting from an initial grid (a) and using a contour prior (b). The linear path $\mathbf{P}_p^k$ is defined for a pixel p and a superpixel S_k (c), and the maximum contour intensity (yellow pixel) is considered to prevent the crossing of image structures (d).

the maximum of contour intensity on $\mathbf{P}_p^k$:

$$d_C(\mathbf{P}_p^k) = 1 + \gamma \max_{q \in \mathbf{P}_p^k} C(q), \tag{7}$$

with $\gamma \geq 0$. Figure 7 illustrates the selection of maximum contour intensity on the linear path. When a high contour intensity is found on the path between a pixel p and the barycenter of S_k, such term prevents this pixel to be associated to the superpixel, and all superpixel boundaries will follow more accurately the image contours. The proposed framework can consider either soft contour maps, *i.e.*, maps having values between 0 and 1, or binary maps. It also adapts well to thick contour prior since only the maximum intensity on the path is considered.

Finally, we multiply this term to the color and spatial distances to ensure the respect of the images contours, and the proposed distance D to minimize during the decomposition is defined as:

$$D(p, C_k) = \left(d_c(p, C_k, \mathbf{P}_p^k) + d_s(p, C_k) \frac{m^2}{r^2} \right) d_C(\mathbf{P}_p^k), \tag{8}$$

with the spatial distance d_s computed as Eq. (1). The SCALP method is summarized in Algorithm 1.

Algorithm 1 SCALP(I, K, C)

Initialization of clusters $C_k \leftarrow [F_k, X_k]$ from a regular grid
Initialization of superpixel labels $\mathcal{S} \leftarrow 0$
Pre-computation of features ${\mathcal{F}_p}^{(2)}$ and ${\mathcal{F}_p}^{(1)}$ (5)
for *each iteration* **do**
    Distance $d \leftarrow \infty$
    for *each* C_k **do**
        for *each* p *in a* $(2r+1)\times(2r+1)$ *pixels window centered on* X_k **do**
            Compute the linear path $\mathbf{P}_p^k$ (Bresenham, 1965)
            Compute $D(p, C_k)$ using C and $\mathbf{P}_p^k$ with (8)
            if $D(p, C_k) < d(p)$ **then**
                $d(p) \leftarrow D(p, C_k)$
                $\mathcal{S}(p) \leftarrow k$
    for *each* C_k **do**
        Update $[F_k, X_k]$
return $\mathcal{S}$

2.4. Initialization Constraint from Contour Prior

In this section, we propose a framework to use an initial segmentation computed from a contour prior completion to constrain the superpixel decomposition. To generate an image segmentation into regions from a contour map requires additional steps but may help to improve the decomposition accuracy. As stated in the introduction, although methods such as Arbelaez and Cohen (2008); Arbelaez et al. (2009) enable to segment an image into partitions considering a contour map, they do not allow to control the size, the shape and the number of the produced regions. We here propose a framework that uses an initial segmentation and produces a regular superpixel decomposition within pre-segmented regions, with control on the number of elements. This way, we take advantage of the initial segmentation accuracy while providing an image decomposition into superpixels of regular sizes and shapes. By initializing the decomposition within the computed regions, the initial superpixels better fit to the image content. For instance, small regions can be initially segmented into one or several superpixels, while they may fall between two initial superpixel barycenters, and would not be accurately segmented during the decomposition process.

2.4.1. Hierarchical Segmentation from Contour Detection

In order to adapt an initial segmentation to produce regular superpixels, we propose to use a hierarchical segmentation, that

can be computed from a contour map with methods such as Arbelaez and Cohen (2008); Arbelaez et al. (2009).

Let $\mathcal{U}$ be a hierarchical segmentation that defines a contour probability map. For any threshold, $\mathcal{U}$ produces a set of closed curves. Regions segmented with low probability, *i.e.*, with low intensity contours in $\mathcal{U}$ can be deleted with a thresholding step. The thresholded closed contour map is denoted $\mathcal{U}_\tau$, for a threshold τ, and its corresponding decomposition into regions is denoted $\mathcal{R}_\tau = \{R_i\}$. Figure 8, illustrates the result obtained from a hierarchical segmentation for several thresholds.

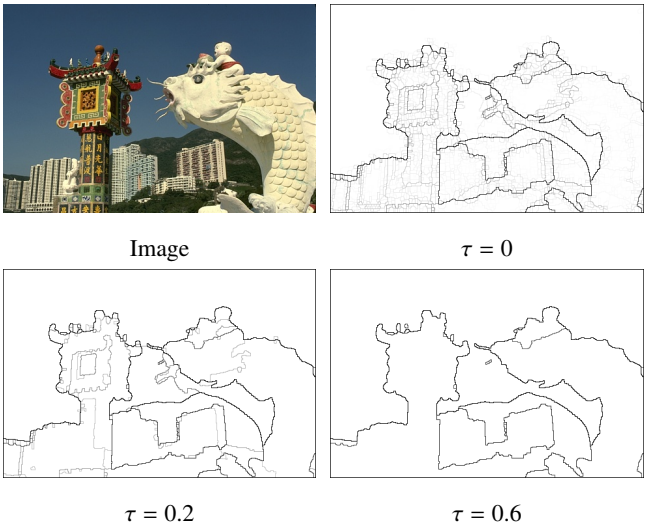


Fig. 8. Example of hierarchical segmentation computed with Arbelaez et al. (2009) from a contour map obtained with Dollár and Zitnick (2013). The hierarchical segmentation is illustrated for several values of the threshold parameter τ.

2.4.2. Regular Decomposition into Superpixels from a Hierarchical Segmentation

Once the hierarchical segmentation is obtained and thresholded, a merging step can be performed to remove the smallest areas. Such small regions should be merged to an adjacent one to respect the size regularity of the decomposition. With K the number of superpixels and $|I|$ the number of pixels of an image I, the superpixel average size is $s = |I|/K$. A threshold $t \in [0, 1]$ is set to merge regions containing less pixels than $s{\times}t$. The segmentation probability of a region R_i is $\min_{p\in\mathcal{B}(R_i)} \mathcal{U}_\tau(p)$, *i.e.*, the lowest intensity among its boundary pixels $p \in \mathcal{B}(R_i)$. The region R_i is hence merged to its adjacent region R_j that shares the boundary with the lowest segmentation probability:

$$\text{if} \quad |R_i| < s{\times}t, \qquad \mathcal{R}_\tau(R_i) = \underset{j, p\in\mathcal{B}(R_i)\cap\mathcal{B}(R_j)}{\operatorname{argmin}} \mathcal{U}_\tau(p). \tag{9}$$

These steps are illustrated in Figure 9, where the thresholding removes areas segmented with low probability and the merging prevents the segmentation of small regions.

A partition step then adds initial superpixels in the remaining regions. If the resulting number of regions is lower than the number of superpixels K, superpixels are added according to the region size $|R_i|$. In a region R_i, $\lfloor |R_i|/s \rfloor$ sub-regions are initialized by a spatial K-means approach (Lloyd, 1982), regardless of the color information.

The proposed approach thus adapts well to the superpixel size, and is not sensitive to threshold settings. The framework using the contour prior as a hard constraint is illustrated in Figure 10, and will be denoted SCALP+HC in the following. Note that although we here consider the segmentation as a hard constraint to enforce the respect of image objects, the image partition can be used to only initialize the superpixel repartition, instead of using a regular grid.

3. Results

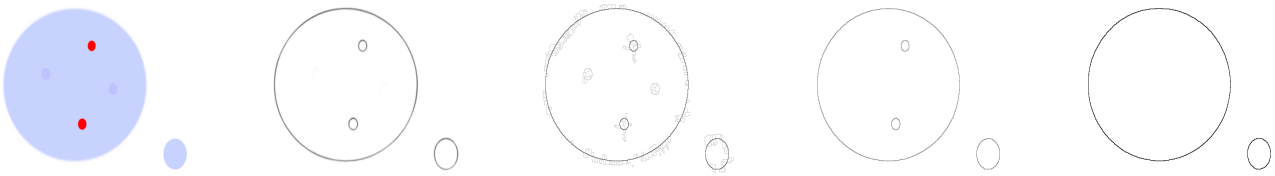
3.1. Validation Framework

3.1.1. Dataset

We use the standard Berkeley segmentation dataset (BSD) (Martin et al., 2001) to evaluate our method and compare to state-of-the-art ones. This dataset contains 200 various test images of size 321×481 pixels. At least 5 human ground truth decompositions are provided per image to compute evaluation metrics in terms of consistency to the image objects, and contour adherence.

3.1.2. Metrics

To evaluate our method and compare to other state-of-the-art frameworks, we use standard superpixel evaluation metrics. The achievable segmentation accuracy (ASA) measures the consistency of the decomposition to the image objects. Boundary recall (BR) and contour density (CD) are used to



(a) Image (b) Contour map (c) Hierarchical segmentation (d) Thresholding (e) Merging

Fig. 9. Illustration of the thresholding and merging steps of the hierarchical segmentation (c) computed from the contour map (b) of an image (a). The thresholding step (d) enables to remove the areas segmented with low probability, *i.e.*, the small blue circles and the segmentation artifacts. Then, according to the condition in Eq. (9), smallest regions are removed (e), *i.e.*, the red circles, although they have higher segmentation probability than the blue ones.

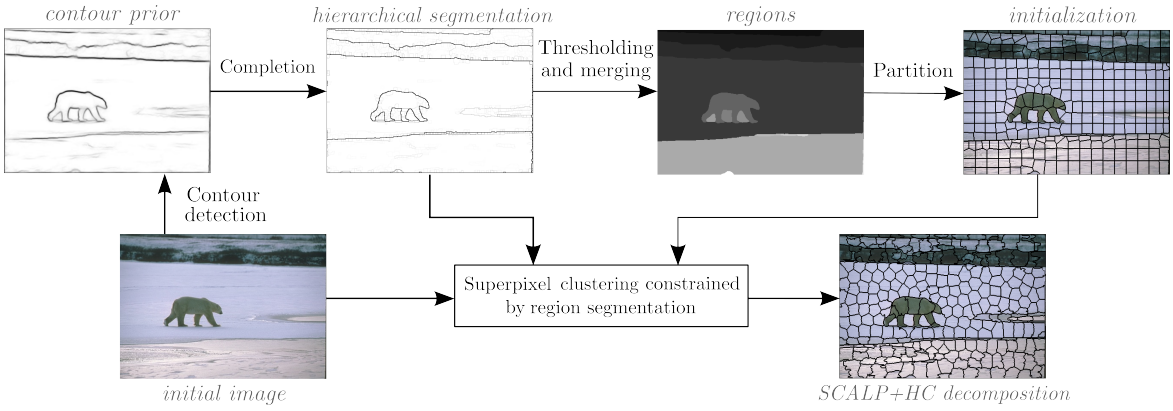


Fig. 10. SCALP+HC framework using the contour prior as a hard constraint to provide an initial segmentation. A completion step produces a hierarchical segmentation from the contour map. Regions segmented with low probability are removed by a thresholding step, and too small regions compared to the given superpixel size are merged to adjacent regions. These regions can then be partitioned to provide a superpixel initialization. SCALP is independently performed in each region, taking advantage of the contour map accuracy while producing a regular decomposition that adapts well to local image content.

measure the detection accuracy according to the ground truth image contours. We also propose to evaluate the contour detection performance of the superpixel methods by computing the precision-recall (PR) curves (Martin et al., 2004). Finally, we report the shape regularity criteria (SRC) (Giraud et al., 2017a) that measures the regularity of the produced superpixels.

For each image I of the dataset, human ground truth segmentations are provided. The reported results are averaged on all segmentations. A ground truth decomposition is denoted $\mathcal{T} = \{T_i\}_{i\in\{1,\dots,|\mathcal{T}|\}}$, with T_i a segmented region, and we consider a superpixel segmentation $\mathcal{S} = \{S_k\}_{k\in\{1,\dots,|\mathcal{S}|\}}$.

Respect of image objects

For each superpixel S_k of the decomposition result, the largest possible overlap with a ground truth region T_i can be computed with ASA, which computes the average overlap percentage for all superpixels:

$$\mathrm{ASA}(\mathcal{S},\mathcal{T}) = \frac{1}{|I|}\sum_k \max_i |S_k \cap T_i|. \tag{10}$$

Note that recent works, *e.g.*, Giraud et al. (2017c); Stutz et al. (2017) show the high correlation between the undersegmentation error (Neubert and Protzel, 2012) and the ASA metric (10). Therefore, the ASA measure is sufficient to evaluate the respect of image objects.

Contour Detection

The BR metric measures the detection of ground truth contours $\mathcal{B}(\mathcal{T})$ by the computed superpixel boundaries $\mathcal{B}(\mathcal{S})$. If a ground truth contour pixel has a decomposition contour pixel at an ϵ-pixel distance, it is considered as detected, and BR is defined as the percentage of detected ground truth contours:

$$\mathrm{BR}(\mathcal{S},\mathcal{T}) = \frac{1}{|\mathcal{B}(\mathcal{T})|}\sum_{p\in\mathcal{B}(\mathcal{T})} \delta[\min_{q\in\mathcal{B}(\mathcal{S})} \|p-q\| < \epsilon], \tag{11}$$

with $\delta[a] = 1$ when a is true and 0 otherwise, and ϵ set to 2 as in, *e.g.*, Van den Bergh et al. (2012). However, this measure only considers true positive detection, and does not consider the number of produced superpixel contours. Therefore, methods that produce very irregular superpixels are likely to have high BR results. To overcome this limitation, as in Machairas et al. (2015); Zhang et al. (2016), the contour density (CD) can be considered to penalize a large number of superpixel boundaries $\mathcal{B}(\mathcal{S})$. In the following, we report CD over BR results, with CD defined as:

$$\mathrm{CD}(\mathcal{S}) = \frac{|\mathcal{B}(\mathcal{S})|}{|I|}. \tag{12}$$

When considering decompositions with the same CD, *i.e.*, the same number of superpixel boundaries, BR results can be relevantly compared. Higher BR with the same CD indicates that the produced superpixels better detect image contours.

The PR framework (Martin et al., 2004) enables to measure the contour detection performances. PR curves consider both boundary recall (BR) (11), *i.e.*, true positive detection, or percentage of detected ground truth contours, and precision $\mathrm{P} = |\mathcal{B}(\mathcal{S}) \cap \mathcal{B}(\mathcal{T})|/|\mathcal{B}(\mathcal{S})|$, *i.e.*, percentage of accurate detection on produced superpixel boundaries. They are computed from an input map, where the intensity in each pixel represents the confidence of being on an image boundary. As in Van den Bergh et al. (2012), we consider the average of superpixel boundaries obtained at different scales, ranging from 25 to 1000 superpixels, to provide a contour detection. In the following, to summarize the contour detection performances, we report the maximum F-measure defined as:

$$\mathrm{F} = \frac{2.\mathrm{P}.\mathrm{BR}}{\mathrm{P} + \mathrm{BR}}. \tag{13}$$

Shape Regularity

To evaluate the regularity of a decomposition in terms of superpixel shape, we use the shape regularity criteria (SRC) introduced in Giraud et al. (2017a), and defined for a decomposition $\mathcal{S}$ as follows:

$$\mathrm{SRC}(\mathcal{S}) = \sum_k \frac{|S_k|}{|I|} . \frac{\mathrm{CC}(H_{S_k})}{\mathrm{CC}(S_k)} \mathrm{V}_{\mathrm{xy}}(S_k), \tag{14}$$

where $\mathrm{V}_{\mathrm{xy}}(S_k) = \min(\sigma_x, \sigma_y)/\max(\sigma_x, \sigma_y)$, evaluates the balanced repartition of the shape S_k with σ_x and σ_y the square root of standard deviations of pixel positions x and y in S_k, H_{S_k} is the convex hull containing S_k, and CC measures the ratio between the perimeter and the area of the considered shape. The SRC measure has been proven to be more robust and accurate than the circularity metric (Schick et al., 2012) used in several superpixel works.

3.1.3. Parameter Settings

SCALP was implemented with MATLAB using single-threaded C-MEX code, on a standard Linux computer. We consider in d_c and d_s more advanced spectral features introduced in Chen et al. (2017). They are designed in a high dimensional space (6 for color, and 4 for spatial features). The linear path between a pixel and the barycenter of a superpixel is computed with Bresenham (1965). In (4), the parameter σ is empirically set to 40 and $\mathcal{P}(p)$ is defined as a 7×7 pixel neighborhood around a pixel p, so $n = 3$. In the proposed color distance (6), λ is set to 0.5, and γ to 50 in (7). The compactness parameter m^2 is set to $0.075r^2$ in the final distance (8), as in Chen et al. (2017). This parameter offers a good trade-off between adherence to contour prior and compactness. The number of clustering iterations is set to 5, contrary to Chen et al. (2017) that uses 20 iterations, since SCALP converges faster. Unless mentioned, when used, the contour prior is computed with Dollár and Zitnick (2013). Finally, when using the contour prior as a hard constraint (SCALP+HC), we respectively set parameters t and τ during the region fusion (9) to 0.15 and 0.4, and compute a hierarchical segmentation with Arbelaez et al. (2009). In the following, when reporting results on noisy images, we use a white additive Gaussian noise of variance 20.

3.2. Influence of Parameters

3.2.1. Distance Parameters

We first measure the influence of the distance parameters in (8) on SCALP performances. In Figure 11, we report results on PR, ASA, CD over BR and SRC curves for different distance settings, on both initial and noisy BSD images. First, we note that the neighboring pixels ($n = 3$ in (4)) increase the decomposition accuracy. The color features ($\lambda = 0.5$ in (6)) also improve

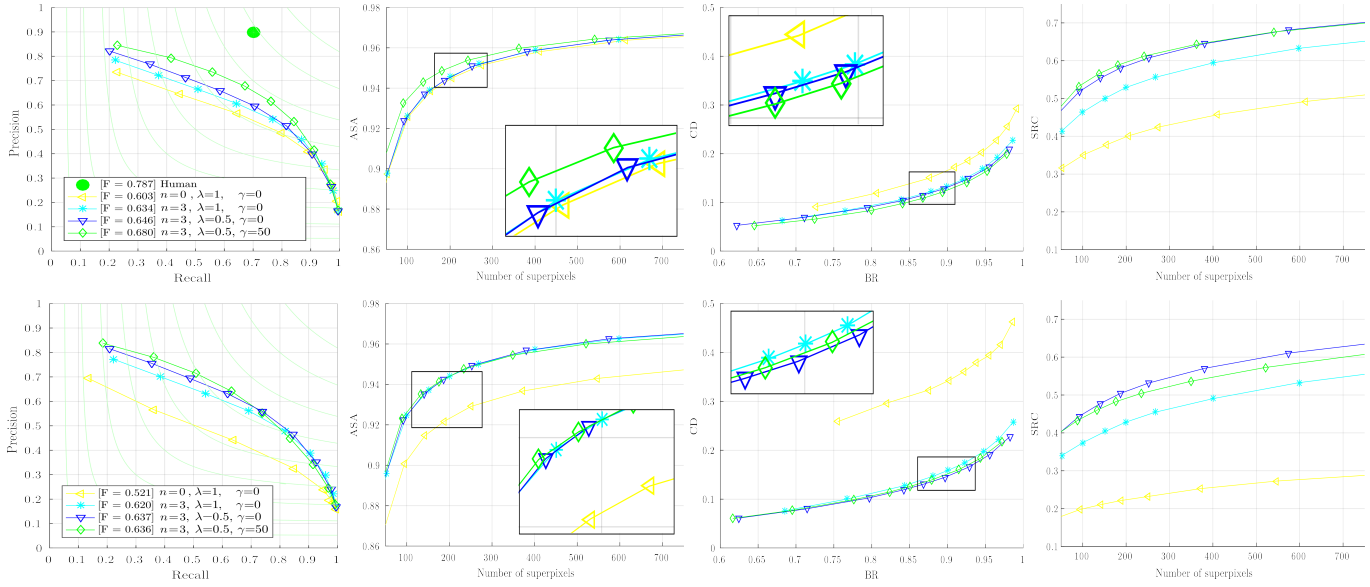


Fig. 11. Evaluation of the SCALP distance parameters on PR, ASA, CD over BR and SRC metrics on initial (top) and noisy images (bottom). Each contribution increases the decomposition accuracy for both initial and noisy images. The parameter n in (4) sets the use of the neighborhood in the clustering distance, and λ in (6) and γ in (7) respectively set the influence of the color distance and contour prior along the linear path. With $n = 0$ in (4), $\lambda = 1$ in (6) and $\gamma = 0$ in (7), the framework is reduced to the method of Chen et al. (2017).

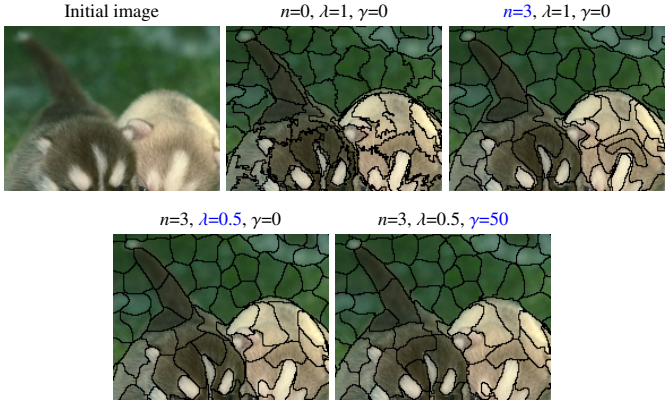


Fig. 12. Visual impact of the distance parameters. Each contribution progressively increases the decomposition accuracy by adding more relevant features.

the results, in terms of respect of image objects and regularity. Finally, the contour prior ($\gamma = 50$ in (7)) along the linear path enables to reach high contour detection (PR) and also increases the performances on superpixel metrics. On noisy images, the accuracy of the contour prior is degraded, but it still provides higher ASA performances on respect of image objects. Note that if $n = 0$ in (4), $\lambda = 1$ in (6) and $\gamma = 0$ in (7), the method is reduced to the framework of Chen et al. (2017).

Figure 12 illustrates the decomposition result for these distance parameters on a BSD image. With only the features used in Chen et al. (2017), *i.e.*, with $n = 0$, $\lambda = 1$, $\gamma = 0$, the decomposition boundaries are very irregular. The neighborhood information greatly reduces the noise at the superpixel boundaries. The color distance on the linear path improves the superpixel regularity and provides more compact shapes. Finally, the contour information enables to more efficiently catch the object structures and to respect the image contours.

3.2.2. Contour Prior

We also investigate the influence of the contour prior. The computation of the contour information should not be sensitive to textures and high local image gradients, and many efficient methods have been proposed in the literature (see for instance references in Arbelaez et al. (2011)). The performances of our method are correlated to the contour detection accuracy, but we demonstrate that improvements are obtained even with basic contour detections.

A fast way to obtain such basic contour detection, which would be robust to textures and high gradients, is to average the boundaries of superpixel decompositions obtained at multiple scales. We propose to consider the same set of scales $\mathcal{K} = \{K\}$

Initial image Boundary average $\bar{\mathcal{B}}$ Contour map C

Fig. 13. Illustration of contour detection from superpixel boundaries computed with Achanta et al. (2012) at multiple scales. Boundaries are averaged and thresholded to provide, in a fast and simple manner, an accurate contour prior. The threshold of the boundary map (15) is set to 0.5.

used for computing the PR curves. All resulting superpixels boundaries $\mathcal{B}(\mathcal{S}^K)$ of a decomposition $\mathcal{S}^K$, computed at scale $K \in \mathcal{K}$ are averaged:

$$\bar{\mathcal{B}} = \frac{1}{|\mathcal{K}|} \sum_{K \in \mathcal{K}} \mathcal{B}(\mathcal{S}^K). \tag{15}$$

The average $\bar{\mathcal{B}}$ can then be thresholded to remove low confidence boundaries and provide an accurate contour prior C. Figure 13 illustrates the computation of the contour prior C from superpixel boundaries. Note that the decompositions at multiple scales K are independent and can be computed in parallel.

In Figure 14, we provide results obtained by using different contour prior: the contour detection from multiple scale decompositions, using Achanta et al. (2012) with a threshold of the boundary map (15) set to 0.5, from the globalized probability of boundary algorithm (Maire et al., 2008), a method using learned sparse codes of patch gradients (Xiaofeng and Bo, 2012), and from a structured forests approach (Dollár and Zitnick, 2013). The results on all metrics are improved with the accuracy of the provided contour detection. Nevertheless, we note that even simple contour priors enable to improve the superpixel decomposition adherence to boundaries. In the following, reported results are computed using Dollár and Zitnick (2013).

3.3. Comparison with State-of-the-Art Methods

We compare the proposed SCALP approach to the following state-of-the-art methods: ERS (Liu et al., 2011), SLIC (Achanta et al., 2012), SEEDS (Van den Bergh et al., 2012), ERGC (Buyssens et al., 2014), Waterpixels (WP) (Machairas et al., 2015), ETPS (Yao et al., 2015) and LSC (Chen et al., 2017). Reported results are computed with codes provided by the authors, in their default settings.

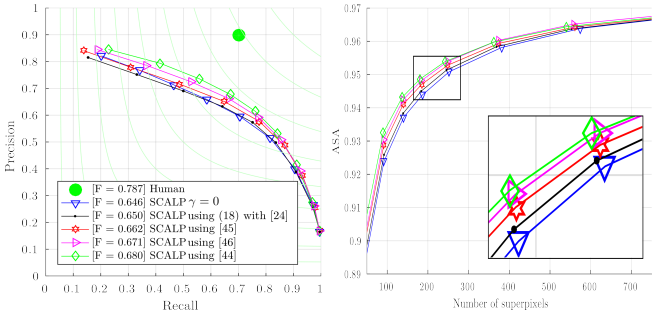


Fig. 14. Evaluation of different contour priors. Even a simple contour detection from averaged superpixel boundaries at multiple scales improves the adherence to image contours.

In Figure 15, we provide PR curves with the maximum F-measure, and report the standard ASA (10), CD (12) over BR (11) and SRC (14) metrics on both initial (top) and noisy (bottom) images. SCALP outperforms the compared methods on the respect of image objects and contour detection metrics, providing for instance higher F-measure (F = 0.680), while producing regular superpixels. The regularity is indeed increased compared to SLIC and LSC, and is among the highest of state-of-the-art methods. The ASA evaluates the consistency of a superpixel decomposition with respect to the image objects, enhancing the largest possible overlap. Therefore, best ASA results obtained with SCALP indicate that the superpixels are better contained in the image objects. Using the contour prior as a hard constraint (SCALP+HC), our method even reaches higher performances, for instance with F = 0.709. Moreover, SCALP results obtained without using a contour prior, *i.e.*, setting γ to 0 in (7), still outperform the ones of the most accurate compared methods LSC and ERS. Finally, we can underline the fact that SCALP results outperform the ones of all the compared state-of-the-art methods on contour detection and respect of image objects metrics while producing regular superpixels. The gain of performances is further assessed by the result of a paired Student test on the ASA result sets. A very low p-value (< 0.002) is obtained by comparing the result set of SCALP to the one of ERS (Liu et al., 2011), the best compared method in terms of accuracy, which demonstrates the significant increase of performances obtained with SCALP. Generally, to enforce the regularity may reduce the contour adherence (Van den Bergh et al.,

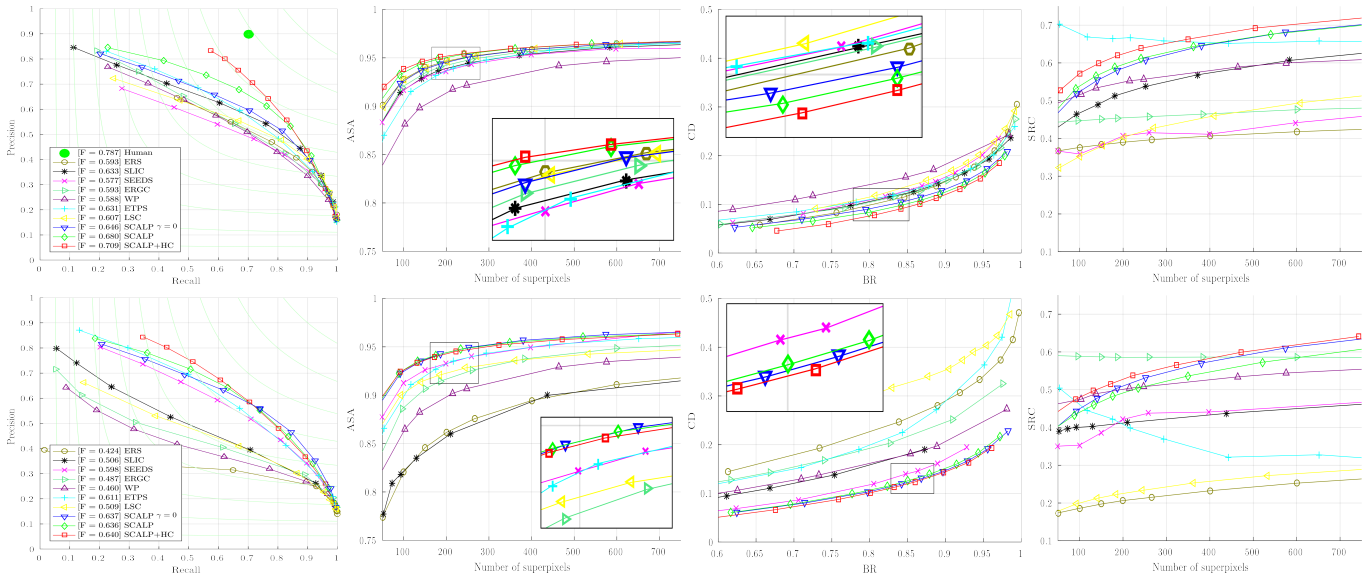


Fig. 15. Comparison between the proposed SCALP framework and the state-of-the-art methods on contour detection (PR) and superpixel metrics (ASA, CD over BR and SRC) on the BSD test set. SCALP outperforms the other methods on both initial images (top) and noisy images (bottom). Moreover the results obtained with SCALP without using a contour prior (SCALP $\gamma = 0$), still outperform the most accurate compared methods (Chen et al., 2017; Liu et al., 2011).

Table 1. Comparison to state-of-the-art methods on initial|noisy images. The maximum F-measure (13) **is computed as described in Section 3.1. ASA** (10) **and SRC** (14) **results are given for $K = 250$ superpixels, and CD** (12) **results for BR** $= 0.8$ (11)**. Blue (bold) and red (underlined) respectively indicate best and second results**

Method	F (13)	ASA (10)	CD/BR (12)	SRC (14)
ERS	0.593|0.424	0.951|0.872	0.099|0.227	0.395|0.213
SLIC	0.633|0.506	0.944|0.867	0.106|0.156	0.537|0.417
SEEDS	0.577|0.598	0.943|0.939	0.109|0.118	0.414|0.435
ERGC	0.593|0.487	0.948|0.924	0.104|0.192	0.457|**0.586**
WP	0.588|0.460	0.932|0.907	0.124|0.162	0.557|0.508
ETPS	0.631|0.509	0.943|0.939	0.110|0.199	**0.663**|0.386
LSC	0.607|0.611	0.950|0.929	0.115|0.300	0.420|0.234
SCALP	0.680|0.636	0.954|**0.949**	0.084|0.107	0.614|0.509
SCALP+HC	**0.709**|**0.640**	**0.955**|0.947	**0.076**|**0.101**	0.641|0.545

2012), but SCALP succeeds in providing regular but accurate superpixels. This regularity property has been proven crucial for object recognition (Gould et al., 2014), tracking (Reso et al., 2013) and segmentation and labeling applications (Strassburg et al., 2015). Therefore, the use of SCALP may increase the accuracy of such superpixel-based methods.

The gain over state-of-the-art methods is largely increased when computing superpixels on noisy images. Methods such as Buyssens et al. (2014); Chen et al. (2017); Liu et al. (2011) obtain very degraded performances when applied to slightly noised images, while Van den Bergh et al. (2012) is the only method that is robust to noise on all evaluated aspects. The state-of-the-art methods can indeed have very different behavior when applied to noisy images. They generally produce very noisy superpixel boundaries (see Figure 1). This aspect is expressed by the lower performances of CD over BR in the bottom part of Figure 15. The regularity is also degraded for all methods, except Buyssens et al. (2014), that tends to generate more regular superpixels, failing at grouping homogeneous pixels. Finally, on the ASA metric, SCALP provides slightly higher results than SCALP+HC for these images. The presence of noise may mislead the contour detection that should not be considered as a hard constraint to ensure the respect of object segmentation. These results are summarized in Table 1, where we report the performances of all compared methods on both initial and noisy images for $K = 200$ superpixels.

Despite the large number of features used in SCALP, the computational time remains reasonable, *i.e.*, less than 0.5s on BSD images, on a single CPU, without any multi-threading architecture, contrary to implementations of methods such as

ETPS (Yao et al., 2015). This computational time corresponds to standard ones of superpixel methods, and SCALP is even faster than methods such as Levinshtein et al. (2009); Liu et al. (2011), whose computational time can be up to 5s.

In this work, we focus on the decomposition performances and do not extensively compare the processing times, since this measurement is highly dependent on the implementation and optimization, and does not necessarily reflect the computational potential of each method (Stutz et al., 2017). Nevertheless, our method is based on the iterative clustering framework (Achanta et al., 2012), and recent works have demonstrated that such algorithm could be implemented to perform in real-time (Ban et al., 2016; Choi and Oh, 2016; Neubert and Protzel, 2014). Therefore, since SCALP have the same complexity as SLIC, our method can reach such computational time with optimized implementation or multi-threading architectures.

Finally, Figures 16 and 17 respectively illustrate the superpixel decomposition results obtained with SCALP and the best compared methods on initial and noisy images. SCALP provides more regular superpixels while tightly following the image contours. SCALP+HC enables to more accurately guide the decomposition by constraining superpixels to previously segmented regions. While most of the compared methods produce inaccurate and irregular results with slightly noised images (see Figure 17), SCALP is robust to noise and produces regular superpixels that adhere well to the image contours.

3.4. Extension to Supervoxels

Finally, we naturally extend the SCALP method to the computation of supervoxels on 3D volumes, for the segmentation of 3D objects or medical images. Many supervoxel methods are dedicated to video segmentation, see for instance Xu and Corso (2012), and references therein. These methods segment the volume into temporal superpixel tubes and are therefore only adapted to the context of video processing. Other methods propose to perform superpixel tracking, *e.g.*, Chang et al. (2013); Reso et al. (2013); Wang et al. (2011), which can result in similar tubular supervoxel segmentation, and may require the computation of optical flow to be efficient (Chang et al., 2013). Contrary to other methods that necessitate substantial adaptations for 3D data, we naturally extend SCALP to compute 3D volume decompositions. We start from a 3D regular grid and perform the decomposition by adding one dimension to the previous equations presented in Section 2.

To validate our extension to supervoxels, we consider 3D magnetic resonance imaging (MRI) data from the Brain Tumor Segmentation (BRATS) dataset (Menze et al., 2015). This dataset contains 80 brain MRI of patients suffering from tumors. The images are segmented into three labels: background, tumor and edema, surrounding the tumor. We illustrate examples of SCALP supervoxel segmentation with the ground truth segmentation in Figure 18, where the tumor and edema are respectively segmented in green and red color. This dataset is particularly challenging since the resolution of images is very low and the ground truth segmentation is not necessarily in line with the image gradients. Finally, note that SCALP obtains an average 3D ASA measure of 0.9848, and outperforms state-of-the-art methods with available implementations SLIC (Achanta et al., 2012) and ERGC (Buyssens et al., 2014), that respectively obtain a 3D ASA of 0.9840 and 0.9652.

4. Conclusion

In this work, we generalize the superpixel clustering framework proposed in Achanta et al. (2012); Giraud et al. (2016), by considering color features and contour intensity on the linear path from the pixel to the superpixel barycenter. Our method is robust to noise and the use of features along such path improves the respect of image objects and the shape regularity of the superpixels. The considered linear path naturally enforces the superpixel convexity while other geodesic distances would provide irregular superpixels. Our fast integration of these features within the framework enables to compute the decomposition in a limited computational time. SCALP obtains state-of-the-art results, outperforming the most recent methods of the literature on superpixel and contour detection metrics. Image processing and computer vision pipelines would benefit from using such regular, yet accurate decompositions.

Fig. 16. Qualitative comparison of decomposition results between SCALP and state-of-the-art superpixel methods on example images of the BSD. SCALP provides the most visually satisfying results with superpixels that adhere well to the image contours while being equally sized and having compact shapes. SCALP+HC enables to further enforce the respect of image contours.

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.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. and Mach. Intell. 34, 2274–2282.

Arbelaez, P., Cohen, L., 2008. Constrained image segmentation from hierarchical boundaries, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–8.

Arbelaez, P., Maire, M., Fowlkes, C., Malik, J., 2009. From contours to regions:

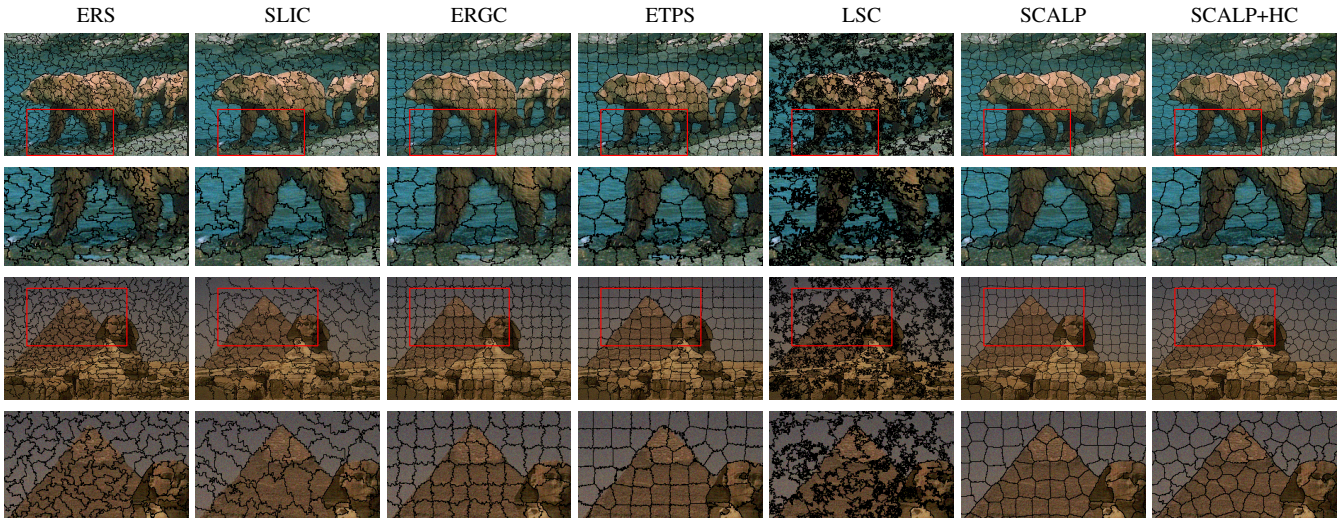


Fig. 17. Qualitative comparison of decomposition results between SCALP and state-of-the-art superpixel methods on noisy images of the BSD. The compared methods generate inaccurate superpixels with noisy borders, while SCALP is robust to noise and produces regular superpixels with smooth boundaries that adhere well to the image contours.

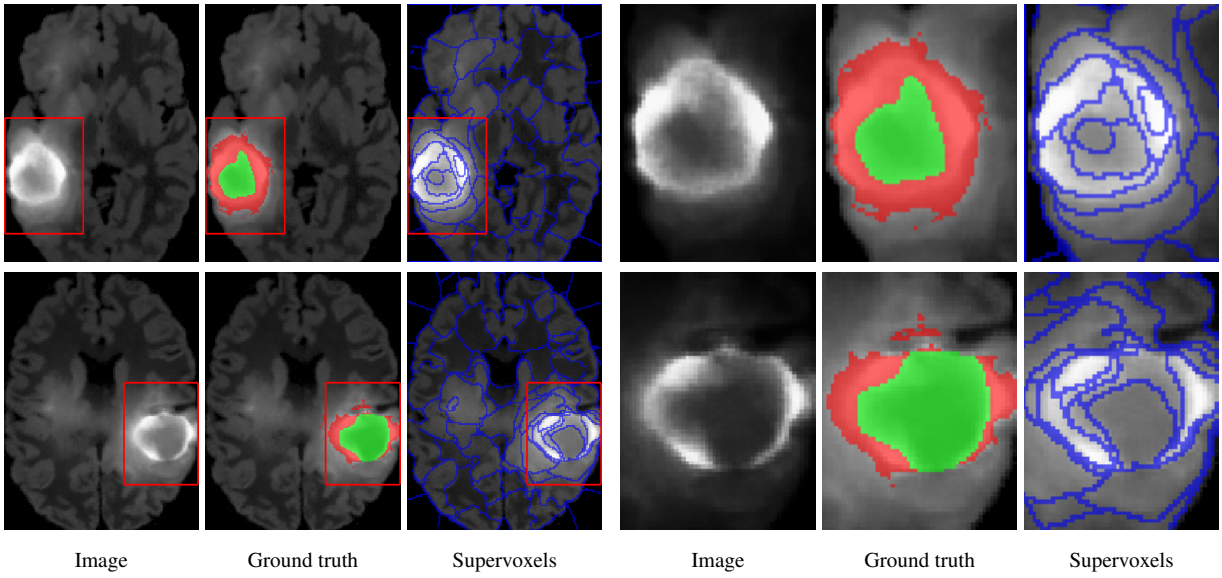


Fig. 18. Examples of supervoxel decomposition with SCALP on medical images (Menze et al., 2015) (left), with a zoom (right) on the tumor and edema, respectively segmented in green and red color. The structures are detected by SCALP despite the low image resolution.

an empirical evaluation, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2294–2301.

Arbelaez, P., Maire, M., Fowlkes, C., Malik, J., 2011. Contour detection and hierarchical image segmentation. IEEE Trans. Pattern Anal. and Mach. Intell. 33, 898–916.

Ban, Z., Liu, J., Fouriaux, J., 2016. GLSC: LSC superpixels at over 130 FPS. J. of Real-Time Image Process. , 1–12.

Bresenham, J.E., 1965. Algorithm for computer control of a digital plotter. IBM Syst. J. 4, 25–30.

Buyssens, P., Gardin, I., Ruan, S., Elmoataz, A., 2014. Eikonal-based region growing for efficient clustering. Image and Vision Computing 32, 1045–1054.

Chang, J., Wei, D., Fisher, J.W., 2013. A video representation using temporal superpixels, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2051–2058.

Chen, J., Li, Z., Huang, B., 2017. Linear spectral clustering superpixel. IEEE Trans. Image Process. (2017) .

Choi, K.S., Oh, K.W., 2016. Subsampling-based acceleration of simple linear iterative clustering for superpixel segmentation. Comput. Vis. Image Underst. 146, 1–8.

Comaniciu, D., Meer, P., 2002. Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. and Mach. Intell. 24, 603–619.

Dollár, P., Zitnick, C.L., 2013. Structured forests for fast edge detection, in: Proc. of IEEE International Conference on Computer Vision, pp. 1841–1848.

Felzenszwalb, P.F., Huttenlocher, D.P., 2004. Efficient graph-based image segmentation. Int. J. Comput. Vis. 59, 167–181.

Fulkerson, B., Vedaldi, A., Soatto, S., 2009. Class segmentation and object localization with superpixel neighborhoods, in: Proc. of IEEE International Conference on Computer Vision, pp. 670–677.

Giraud, R., Ta, V.T., Papadakis, N., 2016. SCALP: Superpixels with contour adherence using linear path, in: Proc. of International Conference on Pattern Recognition, pp. 2374–2379.

Giraud, R., Ta, V.T., Papadakis, N., 2017a. Robust shape regularity criteria for superpixel evaluation, in: Proc. of IEEE International Conference on Image Processing.

Giraud, R., Ta, V.T., Bugeau, A., Coupé, P., Papadakis, N., 2017b. SuperPatchMatch: An algorithm for robust correspondences using superpixel patches. IEEE Trans. Image Process. (in press).

Giraud, R., Ta, V.T., Papadakis, N., 2017c. Evaluation framework of superpixel methods with a global regularity measure. J. of Electronic Imaging 26, 26–18.

Gould, S., Rodgers, J., Cohen, D., Elidan, G., Koller, D., 2008. Multi-class segmentation with relative location prior. Int. J. Comput. Vis. 80, 300–316.

Gould, S., Zhao, J., He, X., Zhang, Y., 2014. Superpixel graph label transfer with learned distance metric, in: Proc. of European Conference on Computer Vision, pp. 632–647.

Gulshan, V., Rother, C., Criminisi, A., Blake, A., Zisserman, A., 2010. Geodesic star convexity for interactive image segmentation, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 3129–3136.

Huang, C.R., Wang, W.A., Lin, S.Y., Lin, Y.Y., 2016. USEQ: Ultra-fast superpixel extraction via quantization, in: International Conference on Pattern Recognition.

Kae, A., Sohn, K., Lee, H., Learned-Miller, E., 2013. Augmenting CRFs with Boltzmann machine shape priors for image labeling, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2019–2026.

Levinshtein, A., Stere, A., Kutulakos, K.N., Fleet, D.J., Dickinson, S.J., Siddiqi, K., 2009. Turbopixels: Fast superpixels using geometric flows. IEEE Trans. Pattern Anal. and Mach. Intell. 31, 2290–2297.

Liu, M.Y., Tuzel, O., Ramalingam, S., Chellappa, R., 2011. Entropy rate superpixel segmentation, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2097–2104.

Lloyd, S., 1982. Least squares quantization in PCM. IEEE Trans. on Information Theory 28, 129–137.

Machairas, V., Faessel, M., Cárdenas-Peña, D., Chabardes, T., Walter, T., Decencière, E., 2015. Waterpixels. IEEE Trans. Image Process. 24, 3707–3716.

Maire, M., Arbelaez, P., Fowlkes, C., Malik, J., 2008. Using contours to detect and localize junctions in natural images, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–8.

Martin, D., Fowlkes, C., Tal, D., Malik, J., 2001. 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, pp. 416–423.

Martin, D., Fowlkes, C., Malik, J., 2004. Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Trans. Pattern Anal. and Mach. Intell. 26, 530–549.

Menze, B., Jakab, A., Bauer, S., et al., 2015. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging 34, 1993–2024.

Moore, A.P., Prince, S.J.D., Warrell, J., Mohammed, U., Jones, G., 2008. Superpixel lattices, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–8.

Mori, G., Ren, X., Efros, A.A., Malik, J., 2004. Recovering human body configurations: combining segmentation and recognition, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 326–333.

Neubert, P., Protzel, P., 2012. Superpixel benchmark and comparison, in: Forum Bildverarbeitung, pp. 1–12.

Neubert, P., Protzel, P., 2014. Compact watershed and preemptive SLIC: on improving trade-offs of superpixel segmentation algorithms, in: Proc. of International Conference on Pattern Recognition, pp. 996–1001.

Reso, M., Jachalsky, J., Rosenhahn, B., Ostermann, J., 2013. Temporally consistent superpixels, in: Proc. of IEEE International Conference on Computer Vision, pp. 385–392.

Rubio, A., Yu, L., Simo-Serra, E., Moreno-Noguer, F., 2016. BASS: Boundary-aware superpixel segmentation, in: Proc. of International Conference on Pattern Recognition.

Sawhney, R., Li, F., Christensen, H.I., 2014. GASP: Geometric association with surface patches, in: Proc. of International Conference on 3D Vision.

Schick, A., Fischer, M., Stiefelhagen, R., 2012. Measuring and evaluating the compactness of superpixels, in: Proc. of International Conference on Pattern Recognition, pp. 930–934.

Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. and Mach. Intell. 22, 888–905.

Strassburg, J., Grzeszick, R., Rothacker, L., Fink, G.A., 2015. On the influence of superpixel methods for image parsing, in: Proc. of the Inf. Conf. on Computer Vision Theory and Applications (VISAPP), pp. 518–527.

Stutz, D., Hermans, A., Leibe, B., 2017. Superpixels: An evaluation of the state-of-the-art. Comput. Vis. Image Underst. (in press).

Tighe, J., Lazebnik, S., 2010. SuperParsing: Scalable nonparametric image parsing with superpixels, in: Proc. of European Conference on Computer Vision, pp. 352–365.

Van den Bergh, M., Boix, X., Roig, G., de Capitani, B., Van Gool, L., 2012. SEEDS: Superpixels extracted via energy-driven sampling, in: Proc. of European Conference on Computer Vision, pp. 13–26.

Vedaldi, A., Soatto, S., 2008. Quick shift and kernel methods for mode seeking, in: Proc. of European Conference on Computer Vision, pp. 705–718.

Veksler, O., Boykov, Y., Mehrani, P., 2010. Superpixels and supervoxels in an energy optimization framework, in: Proc. of European Conference on

Computer Vision, pp. 211–224.

Vincent, L., Soille, P., 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans. Pattern Anal. and Mach. Intell. 13, 583–598.

Wang, P., Zeng, G., Gan, R., Wang, J., Zha, H., 2013. Structure-sensitive superpixels via geodesic distance. Int. J. Comput. Vis. 103, 1–21.

Wang, S., Lu, H., Yang, F., Yang, M.H., 2011. Superpixel tracking, in: Proc. of IEEE International Conference on Computer Vision, pp. 1323–1330.

Xiaofeng, R., Bo, L., 2012. Discriminatively trained sparse code gradients for contour detection, in: Proc. of Conf. on Neural Information Processing Systems, pp. 584–592.

Xu, C., Corso, J.J., 2012. Evaluation of super-voxel methods for early video processing, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1202–1209.

Yang, Y., Hallman, S., Ramanan, D., Fowlkes, C., 2010. Layered object detection for multi-class segmentation, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 3113–3120.

Yao, J., Boben, M., Fidler, S., Urtasun, R., 2015. Real-time coarse-to-fine topologically preserving segmentation, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2947–2955.

Zhang, N., Zhang, L., 2017. SSGD: Superpixels using the shortest gradient distance, in: Proc. of the IEEE International Conference on Image Processing.

Zhang, Y., Li, X., Gao, X., Zhang, C., 2016. A simple algorithm of superpixel segmentation with boundary constraint. IEEE Trans. Circuits and Syst. for Video Technol. PP, 1–1.