



**HAL**  
open science

# Thin Structures Segmentation Using Anisotropic Neighborhoods

Christophe Ribal, Nicolas Lermé, Sylvie Le Hégarat-Mascle

► **To cite this version:**

Christophe Ribal, Nicolas Lermé, Sylvie Le Hégarat-Mascle. Thin Structures Segmentation Using Anisotropic Neighborhoods. 2020. hal-02468105v1

**HAL Id: hal-02468105**

**<https://hal.science/hal-02468105v1>**

Preprint submitted on 5 Feb 2020 (v1), last revised 27 Oct 2020 (v2)

**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.

# Thin Structures Segmentation Using Anisotropic Neighborhoods

Christophe Ribal, Nicolas Lermé, and Sylvie Le Hégarat-Masclé

Université Paris-Saclay, ENS Paris-Saclay, CNRS, SATIE, 94235, Cachan, France  
`{christophe.ribal,nicolas.lerme,sylvie.le-hegarat}@universite-paris-saclay.fr`

**Abstract.** Bayesian and probabilistic models are widely used in image processing to handle noise due to various alteration phenomena. To benefit from the spatial information in a tractable way, Markov Random Fields (MRF) are often assumed with isotropic neighborhoods, that is however at the detrimental of the preservation of thin structures. In this study, we aim at relaxing this assumption on stationarity and isotropy of the neighborhood shape in order to get a prior probability term that is relevant not only within the homogeneous areas but also close to object borders and within thin structures. To tackle the issue of neighborhood shape estimation, we propose to base on tensor voting, that allows for the estimation of structure direction and saliency at various scales. We propose three main ways to derive anisotropic neighborhoods, namely shape-based, target-based and cardinal-based neighborhood. Then, having defined the neighborhood field, we introduce an energy minimized using graph cuts, and illustrate the benefits of our approach against the use of isotropic neighborhoods in the applicative context of crack detection. First results on such a challenging problem are very encouraging.

**Keywords:** Thin structures · Segmentation · Anisotropic neighborhoods · Superpixels · Graph cuts.

## 1 Introduction

Image segmentation is a challenging task in the computer vision field, which deals with the problem of partitioning an image (or video) into multiple regions with labels that may later be used in higher level tasks, like object classification, detection or tracking. This is an ill-posed problem. At pixel level, such operation is prone to noise, corrupted data and all kind of optic phenomena altering the original image. A common way to overcome these difficulties is to take into

account spatial relationships between close pixels in order to favor regular solutions. Classically, one may model the 2-dimensional field of labels as a MRF [10] and compute the segmentation using the Maximum A Posteriori (MAP).

Variational approaches are widely used to provide solutions minimizing a functional gathering energy terms representing data fidelity and regularization terms. The numerous energy models for reducing the impact of image artifacts over the output segmentation nevertheless tend to share a common drawback: They behave poorly on thin structures<sup>1</sup>, because of the small size and complex geometry of the latter with respect to neighborhood ones. The early removal of such structures is a well known effect of Total Variation (TV) regularization (e.g. in image reconstruction [20]) and Potts regularization (e.g. in image segmentation [14]). Thin structures are however ubiquitous in a number of applications (such as medical imaging or quality control) and detecting them as accurately as possible is therefore of great interest.

Alternatively, superpixel decomposition methods have been developed for grouping pixels sharing similar radiometric intensities into regions of controlled spatial extent. Superpixel partitions are generally seen as oversegmentations that preserve small structures but also noise. The benefits of superpixel decomposition is thus to drastically reduce the number of elements to process while keeping the geometrical information that is often lost with multi-resolution approaches and leaving noise removal for further processing steps.

Dealing with further processing, a major drawback of a superpixel segmentation is that the usual hypothesis of a regular lattice is lost (i.e. pixels are all of the same size and shape). As a result, image segmentation approaches taking advantage of superpixels must cope with these problems and introduce new spatial relationships. This induces a neighborhood construction step even for isotropic neighborhoods: For instance, a simple criterion is that superpixels are considered as neighbors when they share a common border ([21,15,9,6]). The authors of [21] propose to minimize an energy using graph cuts on the adjacency graph obtained from the watershed of the input image. In this graph, edges connecting two adjacent regions are weighted upon their common border length, similarly to [6]. Those neighborhood fields based on adjacency do not favor specific orientations of neighborhoods with respect to superpixel context and/or location. The approach of [11] is to gather all superpixels whose centroid belongs to a disc

---

<sup>1</sup> In  $n$ -dimensional images, thin structures or tubular structures are characterized by a significantly smaller size in at least one of their  $n$  dimensions.

centered on it and is therefore isotropic as in most of the other superpixel-based segmentation approaches.

At pixel level however, anisotropic approaches have been introduced to minimize the alteration of thin structures by regularization processes [13,7]. It corresponds to the relaxation of the isotropic hypothesis often introduced when formulating the problem as a MRF. As an alternative to the weighted Total Variation (TV) [19], the authors of [17] introduce a directional TV approach, based on a “vesselness feature” which aims to detect thin structures. Finally, since we believe that structure orientation estimation is a key aspect of anisotropic regularization approaches, let us mention different ways to estimate it: tensor voting approaches [16,5], vesselness operators like RORPO [18], the Frangi vesselness [8], or structure aware regression filters [23] to perform structure-dependent image smoothing. By analogy with usual probabilistic modeling, uniform hypothesis widely used in the absence of prior knowledge corresponds to isotropic neighborhood, and specific prior distribution corresponds to anisotropic neighborhood which can be derived from the observation of the local orientation in our case.

We thus propose a methodology that both allows for the relaxation of isotropic neighborhood which is all the more relevant that we consider superpixel level, and provides regularized results robust to noise. We consider in this context the construction of elliptic neighborhoods, that originate from [7] and [11], and of two path-based neighborhoods. Similarly to [17], we expect these anisotropic neighborhoods to take into account the orientations of image’s structures. Thus, we introduce a new field embedding these orientations computed from tensor voting [16]. Finally, we formulate the segmentation problem in an energy minimization framework, and solve it using graph cuts.

The rest of the paper is organized as follows. The problem formulation is presented in Section 2 and the construction of isotropic and anisotropic neighborhoods is detailed in Section 3. Section 4 introduces the energy terms implemented and Section 5 compares our results against those obtained with isotropic neighborhoods on real and simulated images. Finally, Section 6 outlines the contributions of the paper and discuss future work.

## 2 Problem definition

A superpixel is a group of pixels, defined by their coordinates in the  $n$ -dimensional space,  $n \in \mathbb{N}_{>0}$  ( $n = 2$  in the experiments presented in Section 5). Since each

pixel belongs to one and only one superpixel, the set  $\mathcal{S}$  of  $K \in \mathbb{N}_{>0}$  superpixels is a partition of the original image. Denoting by  $\mathcal{P}$  the set of pixels, then each superpixel  $s$  is an element of  $\mathcal{S} \subset 2^{\mathcal{P}}$ , i.e.  $s \in 2^{\mathcal{P}}$  where  $2^X$  denotes the power set of a set  $X$ . The partition constraint implies that  $\forall p \in \mathcal{P}, \exists! s \in \mathcal{S}$  such that  $p \in s$ . Notice that the shape of any superpixel is also usually constrained to be composed of a single connected component. We denote by  $\mathbb{F}$  the feature space holding the spectral information associated to any pixel or superpixel, for instance  $\mathbb{R}$  (grayscale images),  $\mathbb{R}^3$  (color images) or a higher dimensional space (hyperspectral images).

To stress that our approach can apply indifferently to an image of pixels or of superpixels, we define the position and the features vector of a superpixel  $s \in \mathcal{S}$ . In our case (but other choices could have been done depending on the application), they are the barycenter of the coordinates (in  $n$ -dimensional space) of the pixels that compose  $s$  and the feature barycenter (in  $\mathbb{F}$ ) of these same pixels: The image of superpixels (with superpixels possibly reduced to a single pixel) being denoted  $\mathbf{I} \in \mathbb{F}^{\mathcal{S}}, \forall s \in \mathcal{S}, I(s) \in \mathbb{F}$  is the mean of color (or spectral) features of all pixels in  $s$ . Given a finite set  $\mathcal{C} = \{1, \dots, C\}$  of  $C \in \mathbb{N}_{>0}$  classes, segmenting the image is equivalent to finding a field of labels  $\mathbf{u} \in \mathcal{C}^{\mathcal{S}}$ .

We use the MAP criterion to assign, given the image  $\mathbf{I}$ , a label  $u_s \in \mathcal{C}$  to  $s, \forall s \in \mathcal{S}$ . To this end, we set up a functional  $E$ , to be minimized over the field of labels  $\mathbf{u} \in \mathcal{C}^{\mathcal{S}}$ , that encompasses different priors on the labeling  $\mathbf{u}$ :

$$E(\mathbf{u}, V) = E_1(\mathbf{u}) + \alpha E_2(\mathbf{u}, V), \quad (1)$$

where  $\alpha \in \mathbb{R}_{>0}$  is a parameter controlling the balance between the data fidelity term  $E_1$  and the smoothness term  $E_2$ , and  $V : \mathcal{S} \rightarrow 2^{\mathcal{S}}$  ( $2^{\mathcal{S}}$  denotes the powerset of  $\mathcal{S}$ ) is the neighborhood field that is fixed. Note that  $E_1$  only depends on the image data and on  $\mathbf{u}$ . Smoothness prior on the labeling  $\mathbf{u}$  yields the smoothness term  $E_2$  that is itself based on neighborhood field definition. In this study, we only consider the second order cliques, and we denote by  $\mathcal{N} \subset \mathcal{S}^2$  the set of second order cliques of superpixels. Note that using such a definition, the superpixels of any pair  $(s, t) \in \mathcal{N}$  are not required to have a common boundary. For any superpixel  $s \in \mathcal{S}$ , we define the neighborhood  $V(s)$  of  $s$  as

$$V(s) = \{t \in \mathcal{S} \mid (s, t) \in \mathcal{N}\}.$$

With the relaxation of isotropy and stationarity constraints on  $V$  comes the need to introduce additional priors. First, we formulate the hypothesis that the structure of the neighborhood of a superpixel depends on the structure of

the objects in the image. We model such information of structure through a symmetric second order tensor field  $\mathcal{T} \in (\mathbb{R}^{n \times n})^{\mathcal{S}}$  and assume that it can be built from the field of labels  $\mathbf{u}$ . Other considered priors yield different ways to construct the neighborhood field  $V$ , presented in the next section.

### 3 Proposed neighborhoods construction

In this section, we aim at defining the neighborhood field  $V : \mathcal{S} \rightarrow 2^{\mathcal{S}}$ , possibly anisotropic and non stationary. To each site  $s \in \mathcal{S}$ , we associate a set of sites  $V(s) \in 2^{\mathcal{S}}$ , where  $s$  is either a superpixel or a pixel, such definition being consistent with  $\mathcal{S} = \mathcal{P}$ . To underline the genericity of our formulation, we consider both cases in our experiments.

The construction of our anisotropic neighborhoods aims at encouraging strong relationships between sites aligned with respect to the directions of the thin structures of the image. As explained in Section 1, the characteristics thereby depicted for these structures, namely orientation and saliency, may be retrieved from vesselness operators ([23,8,18]) or Tensor Voting ([16]). In this study, we consider this latter approach where a scale parameter  $\sigma \in \mathbb{R}_{>0}$  sets the span of the voting field. Whatever the way they have been estimated, let us represent the thin structure features in a field of second order tensors  $\mathcal{T} \in (\mathbb{R}^{n \times n})^{\mathcal{S}}$ . For any site  $s \in \mathcal{S}$ , local orientation and saliency of structure are derived from the eigenvectors and the eigenvalues of the tensor  $\mathcal{T}_s$ . Eigenvectors are ranked by decreasing order of their corresponding eigenvalue. More precisely, for any site  $s \in \mathcal{S}$ , the construction of  $V$  is achieved using a set of vectors,  $(\vec{v}_i(\mathcal{T}_s))_{i=0}^{n-1}$ , where  $\vec{v}_i(\mathcal{T}_s) \in \mathbb{R}^n$  is collinear with the  $i^{th}$  eigenvector with its norm being equal to the  $i^{th}$  eigenvalue,  $\forall i \in \{0, \dots, n-1\}$ .

We distinguish two families of anisotropic neighborhoods, namely shape-based neighborhoods and path-based neighborhoods, both compared (see Section 5) against the following neighbourhoods: Stawiaski's boundary-based neighborhood [21] and Giraud's neighborhood [11]. Note that the latter can be seen as an isotropic restriction of our shape-based neighborhood. Path-based neighborhoods stem from the idea of adapting the neighborhood structure to 1-dimensional thin structures, represented by paths. Formally, for any  $k \in \mathbb{N}_{>0}$ , we define a path of cardinality  $k$  as a set of sites  $(s_1, \dots, s_k) \in \mathcal{S}^k$  such that, in our case,  $(s_i, s_{i+1})$  have a common boundary (see [21]),  $\forall i \in \{1, \dots, k-1\}$ . Moreover, we denote by  $\mathcal{P}_K$  the set of paths of cardinality  $K$  for any  $K \in \mathbb{N}_{>0}$  and by  $\mathcal{P}$  the set of paths of any cardinality  $k \in \mathbb{N}_{>0}$ . In what follows, we detail

three different ways to construct the neighborhood field  $V$  using the tensor field  $\mathcal{T}$ .

### 3.1 Shape-based neighborhoods

We are inspired of [11], using superpixel centroid relative locations and  $n$ -dimensional shapes instead of discs to settle the shape-based neighborhoods (**shape**). Whenever the centroid of a superpixel belongs to the computed neighborhood shape of a second one, it is added to the neighborhood of the latter. For computational reasons, we discretize the orientations of the parametric shapes based on the one of  $\vec{v}_0(\mathcal{T}_s)$  for any site  $s \in \mathcal{S}$ , which boils down to the use of a dictionary of neighborhood shapes. Notice that the neighborhood  $V(s)$  of any site  $s \in \mathcal{S}$  is not necessarily connected with such an approach.

### 3.2 Target-based neighborhoods

Target-based neighborhood (**target**) is a path-based neighborhood that aims at constructing the neighborhood  $V(s)$  of a site  $s \in \mathcal{S}$  by connecting it to two distant sites  $t_0, t_1 \in \mathcal{S}$  (named “target”) through paths of minimal energy. Hence, the connectedness along these paths (and so  $V(s)$ ) is thus ensured by definition. We propose to find these paths in two stages. Firstly, for any  $j \in \{0, 1\}$ , targets connecting  $s$  are found with

$$t_j^* \in \underset{t \in \tilde{V}(s)}{\operatorname{argmin}} \|I(s) - I(t)\|_2^2 - \beta \|\vec{st}\|^2 \operatorname{sign}((-1)^j \langle \vec{v}_0(\mathcal{T}_s), \vec{st} \rangle), \quad (2)$$

where  $\beta \in \mathbb{R}_{>0}$  is a free parameter,  $\tilde{V}(s)$  denotes a **shape**-based neighborhood (see Section 3.1),  $\operatorname{sign}(\cdot)$  denotes the sign of a real number,  $\langle \cdot, \cdot \rangle$  denotes the scalar product,  $\|\cdot\|$  denotes the Euclidean norm and  $\vec{st}$  denotes the vector connecting any pair of sites  $(s, t) \in \mathcal{S}^2$ . In Eq. (2), the first term favors the sites  $s$  and  $t$  to have similar image intensities while the second term favors far targets from  $s$  that are aligned with  $\vec{v}_0(\mathcal{T}_s)$ .

Secondly, paths of minimal energy connecting the site  $s \in \mathcal{S}$  to either targets  $t_0^*$  or  $t_1^*$  (see Eq. (2)) are obtained with

$$p_j^* \in \underset{p=(s_1=s, \dots, s_{\#p}=t_j^*) \in \mathcal{P}}{\operatorname{argmin}} \sum_{i=1}^{\#p-1} \|I(s_i) - I(s_{i+1})\|_2^2, \quad (3)$$

where  $\sharp$  stands for the cardinality of a set. The term to be minimized in Eq. (3) is large when image intensities of successive sites along a path are dissimilar and small otherwise. Finally, the neighborhood  $V(s)$  of the site  $s$  can be now constructed as follows:  $V(s) = (p_0^* \cup p_1^*) \setminus \{s\}$ .

### 3.3 Cardinal-based neighborhoods

Cardinal-based neighborhood (**cardinal**) is a path-based neighborhood that aims at constructing the neighborhood  $V(s)$  of a site  $s \in \mathcal{S}$  by finding two paths of minimal energy starting from  $s$ , in opposite directions (according to  $\vec{v}_0(\mathcal{T}(s))$  and of fixed length  $K \in \mathbb{N}_{>0}$ . For any  $j \in \{0, 1\}$ , these paths are obtained with

$$p_j^* \in \underset{p=(s_1=s, \dots, s_K) \in \mathcal{P}_K}{\operatorname{argmin}} l_C^j(p). \quad (4)$$

In the above expression,  $l_C^j(p)$  provides a measure of the length of the path  $p$  starting from  $s$ . For any path  $p = (s_1 = s, \dots, s_K) \in \mathcal{P}_K$ ,  $l_C^j(p)$  is defined by

$$l_C^j(p) = \sum_{i=2}^K \|I(s) - I(s_i)\|_2^2 + \beta' \phi_j(\vec{v}_0(\mathcal{T}_s), \vec{s}s_i), \quad (5)$$

where  $\beta' \in \mathbb{R}_{>0}$  is a free parameter and

$$\phi_j(\vec{u}, \vec{v}) = \begin{cases} \arccos \left( \left| \frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\| \|\vec{v}\|} \right| \right) & \text{if } (-1)^j \langle \vec{u}, \vec{v} \rangle > 0, \\ +\infty & \text{otherwise,} \end{cases}$$

measures the angle between the vectors  $\vec{u}$  and  $\vec{v}$  and discriminates whether the scalar product between them is positive or not. The first term of Eq. (5) encourages the image intensities of any site  $s_i$  to be similar to  $s$  while the second term aims at aligning the path with  $\vec{v}_0(\mathcal{T}_s)$ . This allows for ensuring that two paths in opposite directions are selected to establish the neighborhood  $V(s)$  of  $s$ . Finally, the neighborhood  $V(s)$  of the site  $s$  can be now constructed as follows:  $V(s) = (p_0^* \cup p_1^*) \setminus \{s\}$ . In the next section, the segmentation model using anisotropic and isotropic neighborhoods is detailed.

## 4 Proposed model

The data fidelity term  $E_1(\mathbf{u})$  in the functional  $E(\mathbf{u}, V)$  (see Eq. (1)) is the energy term derived from the likelihood  $\mathbb{P}(I | \mathbf{u})$ . At pixel level, popular models



rely on statistical assumptions, especially by assuming site conditional independence. Relying on the same assumption but at superpixel level, the probability  $\mathbb{P}(I | \mathbf{u})$  is the product, over  $\mathcal{S}$ , of probabilities  $\mathbb{P}(I(s) | u_s)$ ,  $I(s) \in \mathbb{F}$  being the observation and  $u_s \in \mathcal{C}$  the class of  $s$ .

In this study, we adopt a color model assuming that image intensities are Gaussian-distributed for each class  $c \in \mathcal{C}$  with mean value  $\mu_c \in \mathbb{F}$  and standard deviation  $\sigma_c \in \mathbb{R}_{>0}$  [4]. Then, the data term  $E_1^s$  for any superpixel  $s \in \mathcal{S}$  and any label  $u_s \in \mathcal{C}$  is written

$$E_1^s(u_s) = \frac{\|I(s) - \mu_{u_s}\|_2^2}{2\sigma_{u_s}^2} + \log(\sigma_{u_s}),$$

and, for any  $\mathbf{u} \in \mathcal{C}^{\mathcal{S}}$ ,  $E_1$  in Eq. (1) is:

$$E_1(\mathbf{u}) = \sum_{s \in \mathcal{S}} E_1^s(u_s). \quad (6)$$

The energy term  $E_2(\mathbf{u}, V)$  corresponds to the smoothness prior on the label field  $\mathbf{u}$  and requires the definition of a neighborhood field  $V$ , as introduced in Section 2. Then, this neighborhood field being fixed, we assume  $\mathbf{u}$  is an MRF so that a prior probability on  $\mathbf{u}$  can be computed from ‘elementary’ energy terms.

In this study, we adopt the Potts model [25], weighted according to the strength of interaction between neighboring superpixels. The definition of  $E_2(\mathbf{u}, V)$  is thus the following:

$$E_2(\mathbf{u}, V) = \sum_{s \in \mathcal{S}} \sum_{p \in V(s)} W(s, p) \mathbb{1}_{\{u_s \neq u_p\}},$$

where  $\mathbb{1}_{\{a \neq b\}} = \begin{cases} 1 & \text{when } a \neq b, \\ 0 & \text{otherwise.} \end{cases}$  and  $W : \mathcal{N} \rightarrow \mathbb{R}_{>0}$  is a weighting function.

For instance, in our implementation of the neighborhood of [21], the weighting function  $W$  is defined for any pair  $(s, p) \in \mathcal{N}$  as  $W(s, p) = \frac{\partial(s, p)}{\partial(s)} \in ]0, 1]$ , where  $\partial(s, p)$  and  $\partial(s)$  denote the common boundary between  $s$  and  $p$  and the perimeter of  $s$ , respectively. In the other neighborhood fields we compare, the cliques  $\mathcal{N}$  can connect non adjacent superpixels. Thus, we propose to define for any pair  $(s, p) \in \mathcal{N}$  the weighting function  $W$  as  $W(s, p) = \#V(s)^{-1} \in ]0, 1]$ .

The Potts model preserves its properties, in particular submodularity, and the data fidelity term  $E_1$  is convex. Numerous works have proven the efficiency of graph cuts [12,3]. According to [12] and [3] respectively, the energy function

defined in Eq. (1) can be exactly minimized when  $\#\mathcal{C} = 2$  (this is the case in our experiments) and approximately minimized when  $\#\mathcal{C} > 2$ .

Finally, let us remark that the estimation of the neighborhood field  $V$  itself requires a segmentation  $\mathbf{u}$ . In this study, we use the blind segmentation (i.e.  $\alpha = 0$ ).

## 5 Numerical experiments

### 5.1 Application framework and parameter setting

Let us now introduce the data and experiments carried out within our application context, namely crack detection. We aim at segmenting a crack, which is a thin structure over a highly textured and noisy background, for instance some asphalt road or concrete wall as in the cracktree dataset [27]. In this study, in addition to images drawn from this dataset, we consider a simulated image with arbitrary shapes and textured noise as shown in Fig. 1. Images intensities are normalized in  $[0, 1]$ .

A variety of algorithms for generating superpixels exist and exhibit different properties [1,22]. Besides, the requirement of providing a partition of the image into connected sets of pixels, main desirable properties are the preservation of image boundaries, the control of the compactness of superpixels and their number, in addition to computational efficiency of the algorithms. In order to study the benefit of our approach also regardless the superpixel decomposition, we propose a “perfectly shaped” set of superpixels generated from the dilated ground truth. Then, for results derivation using actual superpixels, we require the following properties for superpixels: good compactness to be efficiently modeled by their centroid, regularity in size while at the same time allowing the grouping of crack pixels into thin superpixels. Given those prerequisites, we have considered Extended Topology Preserving Segmentation superpixels (ETPS) [26] after image smoothing with median filtering with a square window of size  $(7 \times 7)$ .

Concerning the construction of anisotropic neighborhoods, the parameters are fixed so that there are 6 neighbors per superpixel in average. For **shape**, this reduces to setting the ellipse’s area to 7 times the mean area of a superpixel, while their flattening is set to 0.6. With **cardinal**, we set  $K = 4$  and  $\beta' = 5 \times 10^{-3}$ . Finally for **target**,  $\beta = 5 \times 10^{-3} \times \beta_R$  where  $\beta_R \in \mathbb{R}_{>0}$  is the radius of the shape-based neighborhood ellipsis.

Finally, we estimate the mean  $\mu_c$  of each class  $c \in \mathcal{C}$  (here,  $\#\mathcal{C} = 2$ ) from the mean of the regions of the ground truth, and we set  $\sigma_c = 1$  for simplicity. To compensate the effect of variation of texture size due to perspective, we set the mean value of classes as an affine function of the vertical position of the superpixel. Future works can handle the parameter estimation in an unsupervised context.

## 5.2 Quantitative and qualitative evaluation

Because our ground truth can be composed of 1 pixel width objects, in order to distinguish between slight mislocation errors and non-detection of some parts of the cracks, we compute the F-measure (FM) at scale  $\epsilon = 2$ , based on the number of true positives (TP), false positives (FP) and false negatives (FN), like in [24,2]. In addition, the crack region and the non-crack area being highly unbalanced (in favor to the non-crack area), we use a high value of  $\gamma = 5$  in FM to increase sensitivity to FN with respect to FP:

$$\text{FM}(\gamma) = \left( 1 + \frac{\gamma^2 \text{FN}}{(1 + \gamma^2) \text{TP}} + \frac{\text{FP}}{(1 + \gamma^2) \text{TP}} \right)^{-1} \in [0, 1]. \quad (7)$$



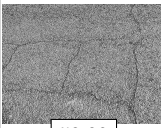
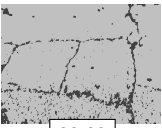
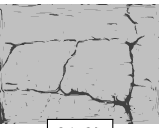

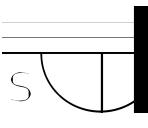
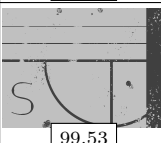
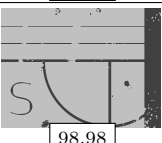
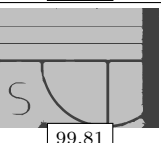
Original image	Ground truth	Blind	Isotropic	Shape
		 58.80	 88.82	 91.97
		 99.53	 98.98	 99.81

Fig. 1: Evaluation performance against ground truths at pixel level for a crack image (top row) and a simulated one (bottom row). The three last columns are segmentations without regularization (“blind”) and with regularization (isotropic with 4-connectivity or **shape**-based anisotropic neighborhoods with ellipses). For each image, in both regularized cases, the results achieving the largest FM with respect to tensor voting scale  $\sigma$  and regularization parameter  $\alpha$ , are depicted. FM measurements are also provided in percents for  $\gamma = 5$ .

The results are presented at pixel level in Fig. 1 and at superpixel level in Fig. 2, respectively. For each image, that corresponds to a set of parameters including the type of superpixels and the type of neighborhood, we select the best result, according to FM criterion, among the results obtained varying the parameters  $\sigma$  (the scale parameter of tensor voting) and  $\alpha$  (the regularization parameter in Eq. (1)).

At pixel level, Fig 1 illustrates the clear improvement of the quality of the results with the use of anisotropic neighborhoods. In the first image of crack, anisotropic regularization allows for enhancing the continuity of the detected cracks, even if some small gaps still fragment it. In the simulated image, the improvement is significant with the correct segmentation of the six discontinuities in the cracks, without loss of precision on more complex shapes.

However, at superpixel level (Fig. 2), while exhibiting better blind results thanks to the averaging of information at pixel level, superpixel anisotropic neighborhoods seem to suffer in general from the fact that it is difficult to establish the right neighborhood  $V$  even with a correct estimation of its orientation (see last column). Our experiments reveal that even if we are far from “Optimal” neighborhood performances, path-based neighborhoods tend to outperform the shape-based ones. Unluckily, the anisotropic approach benefits exhibited in Fig. 2 do not seem to improve the segmentation of the crack image in a so significant way: Path-based approach outperforms the other approaches when superpixels are perfectly shaped, but are still sensitive to the degradation of the quality of the superpixels.

## 6 Conclusion

In this paper, we introduced three anisotropic neighborhoods, in order to make them able to fit the thin structures of the image and thus to improve segmentation results. They rely on the estimation of the orientations of such structures, based here on tensor voting that is efficient in estimating dense map of orientations from a sparse field of labeled sites in the blind segmentation. We then perform the minimization of our energy functional via graph cuts.

We tested our results with a simulated image and an actual difficult crack image, to validate the improvements brought by anisotropic regularization. While our results exhibit a high gain of performances at pixel level, superpixel segmentation suffers from the challenging task to estimate neighborhood at superpixel level, that seems to weaken the benefits of anisotropic regularization.

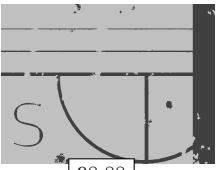
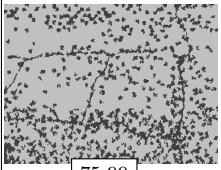
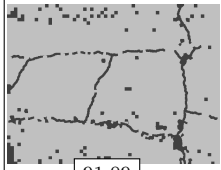
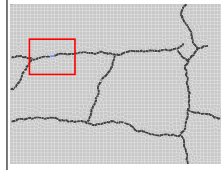

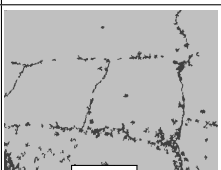

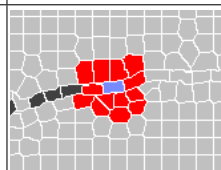

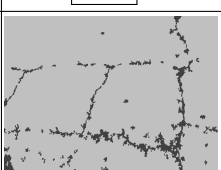
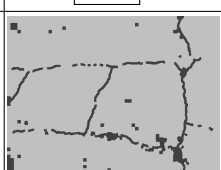
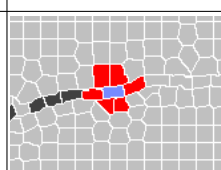
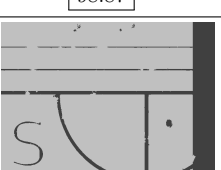
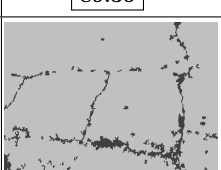
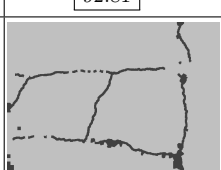
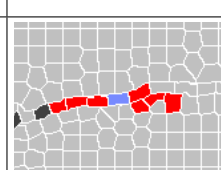
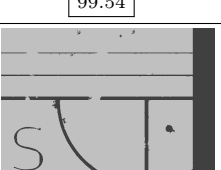
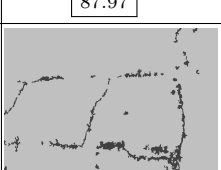
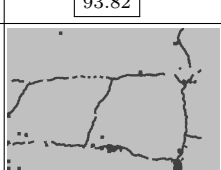
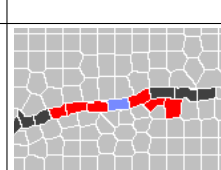
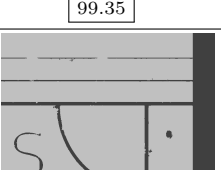
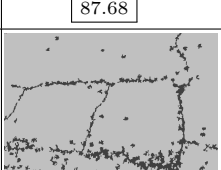
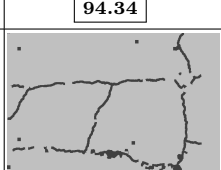
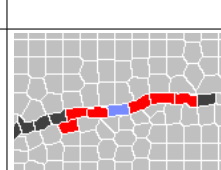
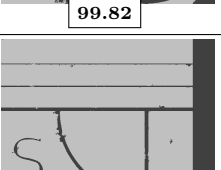
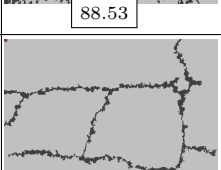
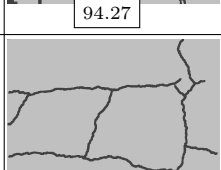
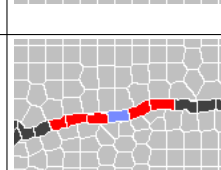
		Simulation		Crack image	
		ETPS [26]		"perfectly shaped"	
Blind					
		98.88	75.99	91.09	
Isotropic	[11]				
		98.84	88.78	92.90	
	[21]				
		98.87	89.36	92.81	
Anisotropic	Shape				
		99.54	87.97	93.82	
	Target				
		99.35	87.68	94.34	
	Cardinal				
		99.82	88.53	94.27	
	Optimal				
		99.94	96.49	98.50	

Fig. 2: Evaluation performance against ground truths at superpixel level for a crack image and a simulated one. The segmentations achieving the largest FM with respect to parameters  $\sigma$  and  $\alpha$  for  $\gamma = 5$ , are depicted. The last two columns correspond to the use of the "perfectly shaped" superpixel, the last one being an illustration of the neighborhood (red) of a superpixel (blue). Each row shows a different type of neighborhood, specified in header lines. The last row is a ground truth-based neighborhood for comparison purpose.

Finally, we plan to investigate the possible refinement of the neighborhood field estimation after computing the regularized segmentation to introduce an alternative minimization procedure, and to explore extensions of our approach with thin structures in shape from focus in 3D-space [20] for future works.

## References

1. Achanta, R., Susstrunk, S.: Superpixels and Polygons Using Simple Non-iterative Clustering. In: Proc. IEEE Conference on CVPR. pp. 4895–4904 (2017)
2. Aldea, E., Le Hégarat-Mascle, S.: Robust crack detection for unmanned aerial vehicles inspection in an *a-contrario* decision framework. J. Electronic Imaging **24**(6), 1–16 (2015)
3. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(11), 1222–1239 (2001)
4. Chan, T., Vese, L.: Active contours without edges. IEEE Transactions on Image Processing **10**(2), 266–277 (2001)
5. Chi-Keung Tang, Medioni, G., Mi-Suen Lee: N-dimensional tensor voting and application to epipolar geometry estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(8), 829–844 (2001)
6. Cui, B., Xie, X., Ma, X., Ren, G., Ma, Y.: Superpixel-Based Extended Random Walker for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing **56**(6), 1–11 (2018)
7. Favaro, P.: Recovering thin structures via nonlocal-means regularization with application to depth from defocus. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 1133–1140 (2010)
8. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A., Delp, S. (eds.) MICCAI, vol. 1496, pp. 130–137 (1998)
9. Fulkerson, B., Vedaldi, A., Soatto, S.: Class segmentation and object localization with superpixel neighborhoods. In: IEEE International Conference on Computer Vision. pp. 670–677 (2009)
10. Geman, S., Geman, D.: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence **6**(6), 721–741 (1984)
11. Giraud, R., Ta, V.T., Bugeau, A., Coupe, P., Papadakis, N.: SuperPatchMatch: An Algorithm for Robust Correspondences Using Superpixel Patches. IEEE Transactions on Image Processing **26**(8), 4068–4078 (2017)
12. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(2), 147–159 (2004)

13. Le Hégarat-Mascle, S., Kallel, A., Descombes, X.: Ant colony optimization for image regularization based on a nonstationary markov modeling. *IEEE Transactions on Image Processing* **16**(3), 865–878 (2007)
14. Lermé, N., Le Hégarat-Mascle, S., Malgouyres, F., Lachaize, M.: Multilayer joint segmentation using MRF and graph cuts. *Journal of Mathematical Imaging and Vision* pp. 1–21 (2020)
15. Liu, Y., Condessa, F., Bioucas-Dias, J.M., Li, J., Du, P., Plaza, A.: Convex Formulation for Multiband Image Classification With Superpixel-Based Spatial Regularization. *IEEE Transactions on Geoscience and Remote Sensing* pp. 1–18 (2018)
16. Medioni, G., Tang, C.K., Lee, M.S.: Tensor Voting: Theory and Applications. In: *Proceedings of RFIA* (2000)
17. Merveille, O., Miraucourt, O., Salmon, S., Passat, N., Talbot, H.: A variational model for thin structure segmentation based on a directional regularization. In: *2016 IEEE International Conference on Image Processing (ICIP)*. pp. 4324–4328 (2016)
18. Merveille, O., Talbot, H., Najman, L., Passat, N.: Curvilinear Structure Analysis by Ranking the Orientation Responses of Path Operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40**(2), 304–317 (2018)
19. Miraucourt, O., Jezierska, A., Talbot, H., Salmon, S., Passat, N.: Variational method combined with Frangi vesselness for tubular object segmentation. In: *Computational & Mathematical Biomedical Engineering (CMBE)*. pp. 485–488 (2015)
20. Ribal, C., Lermé, N., Le Hégarat-Mascle, S.: Efficient graph cut optimization for shape from focus. *Journal of Visual Communication and Image Representation* **55**, 529–539 (2018)
21. Stawiaski, J., Decencière, E.: Region merging via graph-cuts. *Image Analysis & Stereology* **27**(1), 39 (2011)
22. Stutz, D., Hermans, A., Leibe, B.: Superpixels: An evaluation of the state-of-the-art. *Computer Vision and Image Understanding* **166**, 1–27 (2018)
23. Su, Z., Zeng, B., Miao, J., Luo, X., Yin, B., Chen, Q.: Relative reductive structure-aware regression filter. *Journal of Computational and Applied Mathematics* **329**, 244–255 (2018)
24. Vandoni, J., Le Hégarat-Mascle, S., Aldea, E.: Crack detection based on a Marked Point Process model. In: *International Conference on Pattern Recognition (ICPR)*. pp. 3933–3938 (2016)
25. Wu, F.Y.: The Potts model. *Reviews of Modern Physics* **54**(1), 235–268 (1982)
26. Yao, J., Boben, M., Fidler, S., Urtasun, R.: Real-time coarse-to-fine topologically preserving segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2947–2955 (2015)
27. Zou, Q., Cao, Y., Li, Q., Mao, Q., Wang, S.: CrackTree: Automatic crack detection from pavement images. *Pattern Recognition Letters* **33**(3), 227–238 (2012)