LAKE DETECTION WITH SENTINEL-1 DATA USING A GRAB-CUT METHOD AND ITS MULTI-TEMPORAL EXTENSION

Nicolas Gasnier, Loïc Denis, Roger Fjørtoft, Frédéric Liege, Florence Tupin

LAKE DETECTION WITH SENTINEL-1 DATA USING A GRAB-CUT METHOD AND ITS MULTI-TEMPORAL EXTENSION

Nicolas Gasnier[†,*]*, Loïc Denis*[‡]*, Roger Fjørtoft*[*]*, Frédéric Liège*[**]*, Florence Tupin*[†]

[†]LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France
[‡]Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France
[*]CNES, avenue E. Belin, Toulouse, France
[**] C-S Group France, 6 rue Brindejonc des Moulinais, Toulouse, France

ABSTRACT

This paper presents a semi-guided method to detect lakes in Sentinel-1 SAR data. The proposed approach is an adaptation of the grab-cut framework developed in [1]. Starting from a coarse bounding box around the lake, an accurate segmentation is extracted using a Conditional Random Field formalism and a graph-cut based optimization. Then an extension of this approach to process jointly a stack of multi-temporal data is presented. A temporal regularization term is introduced to control the joint segmentation.

The proposed approach is evaluated on Sentinel-1 datasets. Qualitative and quantitative results demonstrate the interest of the proposed framework and its robustness to the initialization polygon of the lake.

***Index Terms*—** SAR imaging, lake detection, grabcut, water surface, Sentinel-1

1. INTRODUCTION

SAR (Synthetic Aperture Radar) is a powerful tool for Earth observation thanks to its all weather and all-time capacities. Among the various applications allowed by such sensors, river and lake monitoring is of the utmost importance, being a key element for water management and climate change monitoring.

The future SWOT mission [2] will give invaluable information thanks to the interferometric potential of KaRIn sensor and its access to height information. Nevertheless, a fine delineation of water bodies could be helpful and could be obtained by complementary information provided by Sentinel-1 data.

In this context, we propose a semi-automatic extraction method of water bodies on sentinel-1 GRD (Ground Range Detected) data. Inspired from the GrabCut method of [1], our approach starts from a coarse bounding box of the water body and extracts an accurate segmentation using a Conditional Random Field (CRF) modeling and a graph-cut based optimization, while iteratively extracting the statistical properties of the water class and the background class thanks to a mixture model. Sentinel-1 provides images of the same area with a temporal frequency of 6 days. To benefit from this situation, the method has been extended to multi-temporal stacks by introducing a temporal regularization.

The proposed framework based on CRF is close to the approaches based on snakes and deforming an initial set of edges or a grid [3][4] which have been adapted to take into account the specificities of the speckle. Besides, many works have been devoted to flood detection using SAR imaging and exploiting the low backscattering coefficient of flooded areas (for example [5]). The context of this work is mainly the use of an available coarse mask of water bodies that could be exploited to derive a finer positioning.

This paper is organized as follows. In section 2 we describe the proposed methodological framework in the case of a single SAR image. This framework is extended to multi-temporal stacks in section 3, whereas section 4 presents some qualitative and quantitative results.

2. PROPOSED METHOD (SINGLE IMAGE CASE)

The objective of the method is to divide the image in two classes: a water class $\mathcal{C}_W$, which corresponds to water pixels (label 1), and a land class which corresponds to land (non-water) pixels (label 0).

The method relies on an initialization step with a rough initial segmentation given by an *a priori* polygon. This polygon could be extracted from an available database. Pixels inside the bounding polygon are given the label $\boldsymbol{\ell}_0 = 1$, while pixels outside the polygon are given the label $\boldsymbol{\ell}_0 = 0$.

Then, the proposed method consists of three steps: the first step assigns each pixel to a class; the second step learns (or updates) the class parameters; the third step updates the segmentation. An overview of the proposed algorithm is presented Fig. 1.

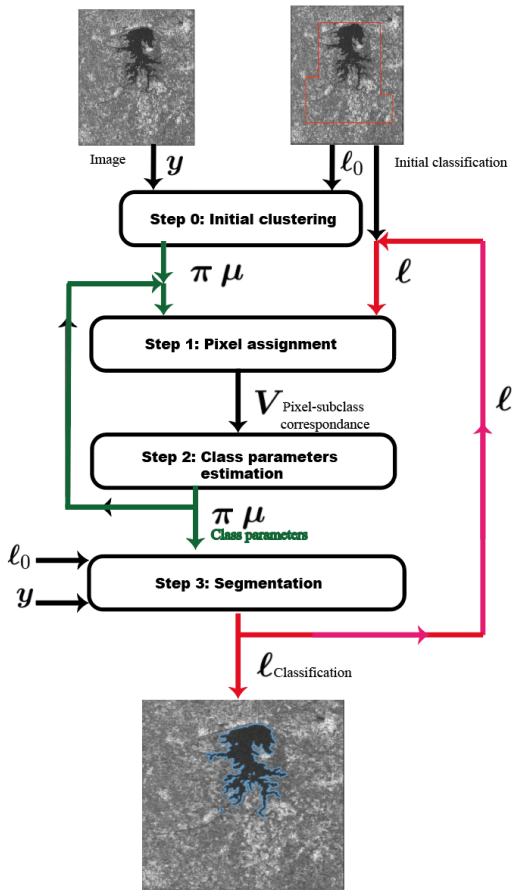

Fig. 1. Synoptic of the proposed approach (notations are detailed in section 2).

2.1. Modeling of the water and land classes and parameter estimation

The method is applied on log-transformed images. In this case the distribution of a physically homogeneous area can be modeled by a Fisher-Tippett distribution. Here, the classes $\mathcal{C}_W$ (water, label 1) and $\mathcal{C}_L$ (land, label 0) are modeled with Fisher-Tippett Mixture Models (FTMM) with respectively $n_{\mathcal{C}_W}$ and $n_{\mathcal{C}_L}$ sub-classes. To take into account land diversity, more classes are used for land: $n_{\mathcal{C}_W} < n_{\mathcal{C}_L}$. Although water is rather homogeneous we chose two sub-classes to take into account wind effects.

Each subclass $\mathcal{K}$ is defined by two parameters that are recomputed at each iteration n_{it}:

- A mean value $\boldsymbol{\mu}(\mathcal{K}, n_{it})$, which is the arithmetic mean of the log-reflectivities of the pixels in the subclass.
- A weight $\boldsymbol{\pi}(\mathcal{K}, n_{it})$ which is the proportion of pixels in $\mathcal{C}$ that belongs to $\mathcal{K}$.

The initialization of these parameters can be done using a K-means approach with the chosen number of sub-classes inside the water and land areas defined by the original polygon. Since the image is log-transformed, the distribution can be approximated by a Gaussian pdf, particularly for multi-looked images.

As opposed to the original GrabCut approach, the variance of the subclass is not considered as a class parameter. Indeed, with the fully developed speckle model with an assumption of homogeneous reflectivity inside of a sub-class, the distribution of log-intensities only depends on the reflectivity and on the equivalent number of looks (ENL) L, which is a characteristic of the image and thus the same for all classes.

The distribution of the log-intensities $\boldsymbol{y}(k)$ assuming a homogeneous reflectivity $R_{\mathcal{K}}$ for all the pixels $k/\boldsymbol{V}_{n_{it}}(k) = \mathcal{K}$ of the subclass $\mathcal{K}$ is given by Fisher-Tippett. With L the ENL of the image and $x_{\mathcal{K}} = \log(R_{\mathcal{K}})$, the log-reflectivity $x_{\mathcal{K}}$ is related to the arithmetic mean of the log-intensities $\boldsymbol{\mu}(\mathcal{K}, n_{it})$ by: $\boldsymbol{\mu}(\mathcal{K}, n_{it}) = x_{\mathcal{K}} - \log(L) + \psi(L)$ and the likelihood for a pixel k of log-intensity $\boldsymbol{y}(k)$ is given by:

$$\mathcal{L}(\boldsymbol{y}(k)|x_{\mathcal{K}}) = \frac{L^L}{\Gamma(L)} e^{L(\boldsymbol{y}(k)-x_{\mathcal{K}})} \exp(-Le^{\boldsymbol{y}(k)-x_{\mathcal{K}}}) \quad (1)$$

2.2. Segmentation step

The goal of the segmentation step is to find a segmentation of the image between the two classes. Each class is modeled by a mixture of Fisher-Tippett distributions defined by the parameters $\boldsymbol{\mu}(\mathcal{K}, n_{it})$ and $\boldsymbol{\pi}(\mathcal{K}, n_{it})$ of its sub-classes as seen in the previous section. The segmentation is obtained by minimizing a global energy $\mathscr{E}$ that depends on the classification $\boldsymbol{\ell}$ ($\boldsymbol{\ell} = 1$ for water and $\boldsymbol{\ell} = 0$ for land). $\mathscr{E}$ is the sum of one data term, a CRF-based regularization term, a flux term and a term preventing water detection outside the *a priori* polygon:

$$\mathscr{E}(\boldsymbol{\ell}, \boldsymbol{y}) = U^I_{\text{data}}(\boldsymbol{\ell}, \boldsymbol{y}) + U_{\text{reg}}(\boldsymbol{\ell}, \boldsymbol{y}) + U_{\text{flux}}(\boldsymbol{\ell}, \boldsymbol{y}) + U_{\text{P}}(\boldsymbol{\ell}, \boldsymbol{\ell}_0) \quad (2)$$

The data term U^I_{data} ensures fidelity to the log-intensity image $\boldsymbol{y} = \log(\boldsymbol{I})$. The regularization term U_{reg} is derived from a CRF model, using a gradient computed with the ROEWA method [6, 7] on the intensities $\boldsymbol{I}$. The term U_{flux} favors a high outward flux of the gradient through the water boundary. An U_{P} term prevents the classification of pixels outside the *a priori* polygon as water.

Data term: In our FTMM, the likelihood of a pixel value given a class $\mathcal{C}$ is the weighted average of the likelihoods of all its sub-classes $\mathcal{K} \in \mathcal{C}$

$$\mathcal{L}_{\boldsymbol{\ell}}(\boldsymbol{y}(k)|\mathcal{C}) = \sum_{\mathcal{K}\in\mathcal{C}} \boldsymbol{\pi}(\mathcal{K}) \mathcal{L}(\boldsymbol{y}(k)|x_{\mathcal{K}})\,. \quad (3)$$

Hence the data term for the whole image is:

$$U^I_{\text{data}}(\boldsymbol{\ell}, \boldsymbol{y}) = -\sum_k \log(\mathcal{L}_{\boldsymbol{\ell}}(\boldsymbol{y}(k)|\boldsymbol{\ell}(k))) \quad (4)$$

Regularization term: A regularization term ensures that the transitions between water and land are compatible with the gradients of the image, by penalizing the transitions that would occur where the gradient magnitude is low, or if the boundaries are not orthogonal to the gradient direction. We want to minimize over the water boundaries the weighted total variation (eq 5) on the label field $\boldsymbol{\ell}$ whose gradient at location u is $\|\overrightarrow{\nabla}\boldsymbol{\ell}(u)\|$. The total variation is weighted by a weight $w_{\text{sym}}(u)$ defined in (6). The regularization energy is then given by equation 5:

$$U_{\text{reg}}(\boldsymbol{\ell}) = \beta \int_{u\in\mathbb{R}^2} w_{\text{sym}}(u)\|\overrightarrow{\nabla}\boldsymbol{\ell}(u)\| \cdot \mathrm{d}u \tag{5}$$

$$\text{with } w_{\text{sym}}(u) = \exp(-|\overrightarrow{\nabla}\boldsymbol{\ell}(u) \cdot \overrightarrow{\nabla}\boldsymbol{I}(u)|/\lambda). \tag{6}$$

This symmetric weighting w_{sym} favors boundaries localizations that are aligned with the edges (strong gradients) of the image. The variables λ and β are parameters that allow adjusting the regularization and its sensitivity to the gradients. This regularization energy can be discretized:

$$U_{\text{reg}}(\boldsymbol{\ell}, \boldsymbol{y}) = \beta \sum_{k'\sim k} w_{\text{sym}}(k,k') \cdot |\boldsymbol{\ell}(k') - \boldsymbol{\ell}(k)| \tag{7}$$

with $w_{\text{sym}}(k,k') = \exp(-|(\boldsymbol{\ell}(k') - \boldsymbol{\ell}(k))(\boldsymbol{I}(k') - \boldsymbol{I}(k))|/\lambda)$, $k' \sim k$ means that k' is an 8-neighbor of k. In the case of pixels that are 8-neighbors of k but not 4-neighbors, λ is multiplied by $\sqrt{2}$. $(\boldsymbol{I}(k') - \boldsymbol{I}(k))$ is actually approximated using the ROEWA gradient computed on the intensity image to smooth out noise.

Flux term: In addition to the data term and the regularization term, a flux term similar to the one proposed in [8] is used to favor or penalize the transitions depending on the orientation and magnitude of the gradient. The gradient is expected to be strong on the water boundary $\partial\{\boldsymbol{\ell} = 1\}$ and oriented in the outward direction. Over the boundary $\partial\{\boldsymbol{\ell} = 1\}$, this criterion locally corresponds to maximizing the dot product between the gradient $\overrightarrow{\nabla}\boldsymbol{I}(u)$ and the unit outward normal vector of the segmentation $\{\boldsymbol{\ell} = 1\}$. Over each entire water body, the criterion can be expressed as the outward flux Φ of the gradient through the boundary $\partial\{\boldsymbol{\ell} = 1\}$ which is equal to the integral of the Laplacian of the image over the water body (divergence theorem). The Laplacian $\overrightarrow{\nabla} \cdot \overrightarrow{\nabla}\boldsymbol{I}(u)$ of the image can be approximated with a Laplacian of Gaussian (LoG) operator of parameter σ_L

$$\overrightarrow{\nabla} \cdot \overrightarrow{\nabla}\boldsymbol{I} \approx \text{LoG}(\boldsymbol{I}, \sigma_L) \tag{8}$$

that can be computed as the convolution of the image $\boldsymbol{I}$ with a precalculated LoG kernel. We call the resulting LoG image $\text{LoG}_{\boldsymbol{y},\sigma_L}$.

The influence of the flux energy $U_{\text{flux}}(\boldsymbol{\ell})$ can be balanced with a scaling parameter $\eta > 0$ that adjusts its effect.

Prior polygon term: A last term U_{P} can be added for improved robustness. It prevents the classification as water of pixels outside of the initial *a priori* polygon (inside this polygon, $\boldsymbol{\ell}_0 = 1$, outside, $\boldsymbol{\ell}_0 = 0$).

$$U_{\text{P}}(\boldsymbol{\ell}, \boldsymbol{\ell}_0) = \sum_{k, \boldsymbol{\ell}(k) > \boldsymbol{\ell}_0(k)} K_C \tag{9}$$

Here, K_C is a parameter.

2.3. Resulting graph and minimization

The resulting global energy can be written as follows after discretization:

$$\begin{aligned}\mathscr{E}(\boldsymbol{\ell}, \boldsymbol{y}) = &\sum_k \log(\mathcal{L}_{\boldsymbol{\ell}}(\boldsymbol{y}(k)|\boldsymbol{\ell}(k))) \\ &+\beta \sum_{k'\sim k} w_{\text{sym}}(k,k') \cdot |\boldsymbol{\ell}(k') - \boldsymbol{\ell}(k)| \\ &+ \sum_k \eta \cdot \text{LoG}_{\boldsymbol{I},\sigma_L}(k) + \boldsymbol{\ell}(k) \cdot (1 - \boldsymbol{\ell}_0(k)) \cdot K_C\end{aligned}$$

where $\overline{\boldsymbol{\ell}}(k) = 1 - \boldsymbol{\ell}(k)$. The global energy can be represented as a cut in a flow network (this is possible as the regularization is sub-modular), and is minimized using the min-cut algorithm proposed in [9].

The resulting partition of the image gives the new labels $\boldsymbol{\ell}(k)$ for each pixel in the image.

3. EXTENSION TO MULTI-TEMPORAL STACKS

The framework is very similar to the one presented in the previous section except that the segmentation is done in 3D and that a temporal regularization term is introduced along the time axis. The segmentation step consists in minimizing a global energy $\mathscr{E}_{\text{MT}}(\boldsymbol{\ell}, \boldsymbol{y})$ that depends on the 2D+T label field $\boldsymbol{\ell}(k,t)$ (k being the spatial index of the pixel and t the date). This global energy is equal to the sum of the energies $\mathscr{E}(\boldsymbol{\ell}(\cdot,t), \boldsymbol{y}(\cdot,t))$ for every image in the stack plus a temporal regularization term between every pair of consecutive images:

$$\begin{aligned}\mathscr{E}_{\text{MT}}(\boldsymbol{\ell}, \boldsymbol{y}) = &\sum_{t=1}^{T} \mathscr{E}(\boldsymbol{\ell}(\cdot,t), \boldsymbol{y}(\cdot,t)) \\ &+ \sum_{t=1}^{T-1} U_{\text{TR}}(\boldsymbol{y}(\cdot,t), \boldsymbol{y}(\cdot,t+1)).\end{aligned} \tag{10}$$

where $\mathscr{E}(\boldsymbol{\ell}(\cdot,t), \boldsymbol{y}(\cdot,t))$ corresponds to the global energy for one image, as defined in equation (7). The temporal regularization $U_{\text{TR}}(\boldsymbol{y}(\cdot,t), \boldsymbol{y}(\cdot,t+1))$ between the images at date t and $t+1$ is the sum of the local regularization energies for every pixels. This regularization energy in (k,t) depends on the difference between the log-intensities $\boldsymbol{y}(k,t)$ and $\boldsymbol{y}(k,t+1)$:

$$\begin{aligned}U_{\text{TR}}((\cdot,t),(\cdot,t+1)) = \beta_T &\sum_k |\boldsymbol{\ell}(k,t) - \boldsymbol{\ell}(k,t+1)| \\ &\exp\left(-\frac{|\boldsymbol{y}(k,t) - \boldsymbol{y}(k,t+1)|}{\lambda_T}\right)\end{aligned} \tag{11}$$

Here, β_T is a tuning parameter for the regularization. To limit the impact of speckle fluctuations, the temporal change $\boldsymbol{y}(k,t)-\boldsymbol{y}(k,t+1)$ can smoothed, for example by convolving with a Gaussian kernel of parameter σ_T, or a better denoising method like [10] [11].

4. RESULTS

In this section, we present the detection results we obtained by applying the proposed approach to a registered stack of 58 Sentinel-1 IW GRD images (geometric product of co- and cross-polarization).

Some qualitative results are displayed in Fig. 2. using a Sentinel-1 GRD image of Sajnam reservoir (India) on 2017-06-03 (above) and 2018-02-22 (below). Right images compare the results to manually-defined ground truths.
For the first date, the F-score is 97.26% vs. 97.03% with its single-date counterpart and 95.79% using a basic MRF method with the true reflectivity for both classes (resp. 95.99%, 95.94% and 94.87% for the second date).
More results can be found in [12] along with a more detailed description of the method. The data and the ground truth used for the experiments are available: https://gitlab.telecom-paris.fr/ring/multitemporal-sar-grabcut.

5. CONCLUSION

The approach presented in this paper is an adaptation of the GrabCut method of [1] to SAR images and the specificities of water detection. A multi-temporal extension has also been presented. The results are satisfying and could be further improved by the use of information on water elevation. The method could be easily used for the updating of water-bodies database.

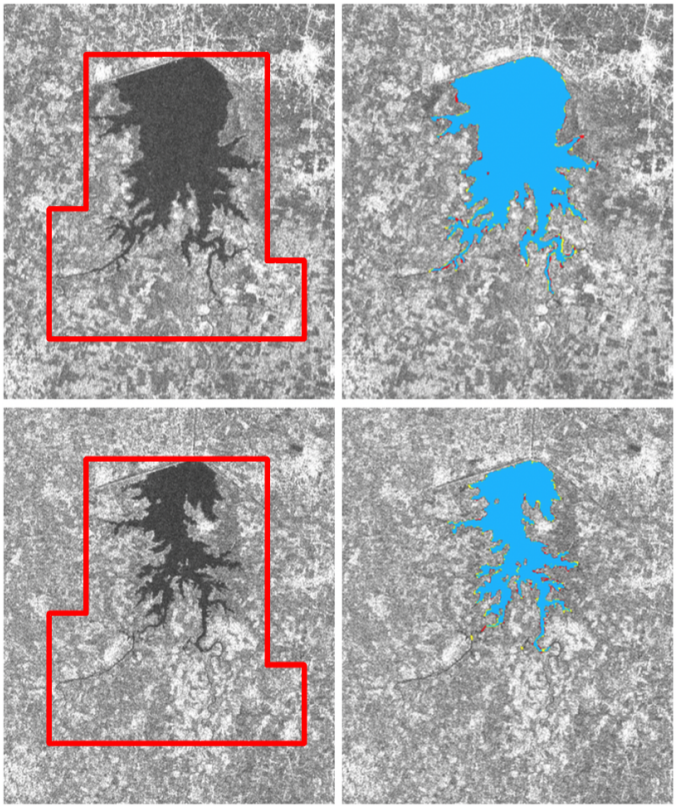

Fig. 2. Initial mask inside the red polygon (left) and segmentation results (rignt) for Sajnam reservoir at two different dates. On the classification errors image (right), the true positives are displayed in blue, the false positives in yellow, and the false negatives in red. True negatives are displayed as the actual SAR image amplitudes.

6. REFERENCES

[1] C. Rother, V. Kolmogorov, and A. Blake, ""GrabCut": Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, p. 309–314, Aug. 2004. [Online]. Available: https://doi.org/10.1145/1015706.1015720

[2] R. Fjortoft and et al., "KaRIn on SWOT: Characteristics of Near-Nadir Ka-Band Interferometric SAR Imagery," *IEEE TGRS*, 2014.

[3] C. Chesnaud, P. Refregier, and V. Boulet, "Statistical region snake-based segmentation adapted to different physical noise models," *IEEE Trans. Pattern Anal. Mach. Intell.*, 1999.

[4] I. B. Ayed, C. Vazquez, A. Mitiche, and Z. Belhadj, "Multiregion level-set partitioning of synthetic aperture radar images," *IEEE Trans. Pattern Anal. Mach. Intell.*, May 2005.

[5] M. Chini, R. Hostache, L. Giustarini, and P. Matgen, "A Hierarchical Split-Based Approach for Parametric Thresholding of SAR Images: Flood Inundation as a Test Case," *IEEE TGRS*, 2017.

[6] R. Fjortoft, A. Lopes, P. Marthon, and E. Cubero-Castan, "An optimal multiedge detector for SAR image segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 793–802, May 1998.

[7] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A SIFT-Like Algorithm for SAR Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 453–466, Jan 2015.

[8] N. Gasnier, L. Denis, R. Fjørtoft, F. Liège, and F. Tupin, "Narrow river extraction from SAR images using exogenous information," *IEEE JS-TARS*, 2021.

[9] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004.

[10] S. Lobry, L. Denis, and F. Tupin, "Multitemporal SAR Image Decomposition into Strong Scatterers, Background, and Speckle," *IEEE JS-TARS*, 2016.

[11] N. Gasnier, E. Dalsasso, L. Denis, and F. Tupin, "Despeckling Sentinel-1 GRD images by deep learning and application to narrow river segmentation," in *IGARSS 2021*, Jul. 2021.

[12] N. Gasnier, "Use of multi-temporal and multi-sensor data for continental water body extraction in the context of the SWOT mission," Theses, Institut Polytechnique de Paris, Jan. 2022. [Online]. Available: https://tel.archives-ouvertes.fr/tel-03578831