Certificate of Attendance

David Alejandro Sierra of Pontificia Universidad Javeriana Cali
[Email: davidjimenez@javerianacali.edu.co] attended IGARSS 2021 and presented the following papers:

- TH4.O·5.2: BLUE NOISE SAMPLING AND NYSTRÖM EXTENSION FOR GRAPH BASED CHANGE DETECTION

IEEE Geoscience and Remote Sensing Society

2021 IEEE International Geoscience and Remote Sensing Symposium
12 – 16 July, 2021
Virtual Event • Brussels, Belgium

IEEE

igarss BRUSSELS 2021
virtual conference • July 12-16, 2021

IEEE GRSS Geoscience and Remote Sensing Society

Michal Shimoni, Devis Tuia, Sindy Stercks and Andrew Skidmore
IGARSS 2021 Technical Program Chairs

BLUE NOISE SAMPLING AND NYSTRÖM EXTENSION FOR GRAPH BASED CHANGE DETECTION

David Alejandro Jimenez-Sierra[†], *Hernán Darío Benítez-Restrepo*[†], *Gonzalo R. Arce*[⋆]
Juan F. Florez-Ospina[⋆]

[†] Pontificia Universidad Javeriana Cali
Departamento de Electrónica y Ciencias de la Computación
{davidjimenez, hbenitez}@javerianacali.edu.co
[⋆] University of Delaware
Multimodal Imaging and Spectroscopy Laboratory
{arce, jfflorez}@udel.edu

IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9555107

ABSTRACT

In this paper, we address the problem of sampling on graphs for change detection in large multi-spectral (MS) and synthetic aperture radar (SAR) images by proposing a graph-based data-driven framework. The main steps of the proposed approach are: (i) the segmentation of regions that enclose the change; (ii) the use of smoothness prior for learning a graph of the regions; (iii) the integration of blue-noise sampling (BN) in the change detection scheme. We validate our approach in 14 real cases of remote sensing according to quantitative analyses. The results confirm that using a structured sampling such as BN outperforms recent state-of-the-art methods in change detection for multimodal data.

Index Terms— Blue-noise, change detection, data fusion, graph, remote sensing images, sampling, smoothness.

1. INTRODUCTION

Change detection (CD) refers to the task of analyzing two or more images acquired over the same area at different times (i.e multitemporal images) to detect land-cover changes between acquisitions [1]. The most common sources are multispectral (MS), and synthetic aperture radar (SAR) images (at very high spatial and spectral resolutions), which describe an object or phenomenon. Each sensor captures different information that explains physical features. For example, a SAR sensor captures information about the physical characteristics of a surface (such as roughness, geometric structure, and orientation), and a MS sensor captures reflectances at different wavelengths from objects [2, 3].

In the last few years, a variety of approaches to CD, based on thresholding, clustering, deep learning, and random Markov fields [4–7], have been proposed. In this work, we focus on graph-based methods due to their capability to handle multimodal data, and the flexibility to represent relationships between data entities [3, 8, 9]. In general, graph-based methods learn a graph from the observed data samples by using a similarity matrix, i.e a Gaussian kernel to quantify the spatial or radiometric proximity between data samples extracted from remote sensing images. Once the graph is learned, the eigenvectors and eigenvalues of the normalized graph Laplacian are used to detect changes in the image. However, when the number of samples is high, computing a complete similarity matrix can be computationally expensive. To address this problem, previous works have proposed sliding windows [8] and Nyström extension (NE) based methods [2, 3]. The sliding windows methods [8] are local approaches that do not take into account the non-local structure present in the image, which may be useful to identify changes. In contrast, NE based methods [2, 3] accounts for non-local structure but overlooks the structural relationships among potential regions of change in the scene. In addition, as it is shown in [10], NE requires as input a set of samples, which highly impacts its approximation capabilities.

Consequently, in this work, we propose to use blue-noise sampling (BN) on a graph [11, 12], which encodes structural relationships of potential regions of change. Then the structurally-aware samples produced by BN are fed into NE. To learn the graph on which BN will be applied, we use graph learning (GL) based on smoothness priors [9, 12, 13], and the potential regions are defined using isoperimetric partitioning [14]. The proposed approach is tested on 14 real datasets in CD and the results are compared with recent state-of-the-art methods such as our previously proposed method based on graphs US-2D [3], U-CD-HPT based on machine learning [8], and rrR based on probabilities distributions [4]. The results show that by using sampling over graphs combined with a GL method to infer the graph, the structural information of the remote sensing image is exploited and outperforms the previous results showed in [3].

2. GRAPH BASED DATA FUSION FOR CHANGE DETECTION

Remote sensing images contain pixels that reside on a regularly sampled 2D grid. Thus, images can be regarded as a signal on a grid graph with edges that connect each pixel in each band to its neighborhood of pixels. Since a graph $G = (V, E, \mathbf{W})$ is a nonlinear structural representation of data, with a set of nodes V, and a set of edges E that explain the relationship among nodes weighted by the entries of the adjacency matrix $\mathbf{W}$. The weight w_{ij} associated with the edge $(i, j) \in E$ quantifies how strong the relationship between the nodes $i, j \in V$ is. The common measure used for each weight is a Gaussian kernel ($\kappa_G(.)$) [3]: $w_{i,j} = \exp\left(-\frac{d(V_i,V_j)^2}{\sigma^2}\right)$, where $d(V_i, V_j)$ is the distance between two nodes and σ is the standard deviation of all $d(V_i, V_j)$. Given the high number of pixels in a remote sensing image (e.g. an image with size 1280×960 is equivalent to $N = 1228800$), the matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$ can be computationally prohibitive to construct. To address this problem, we use the Nyström extension [15] to find an approximation of $\mathbf{W}$, by sampling n_s points distributed across the image and re-indexing the matrix $\mathbf{W}$ as:

$$\mathbf{W} = \kappa_G\left(\begin{bmatrix} \mathbf{d_{AA}} & \mathbf{d_{AB}} \\ \mathbf{d_{AB}}^\top & \mathbf{C} \end{bmatrix}\right),$$

where κ_G is a Gaussian kernel, $\mathbf{d_{AA}} \in \mathbb{R}^{n_s\times n_s}$ represents the graph distances within the n_s sample nodes, $\mathbf{d_{AB}} \in \mathbb{R}^{n_s\times(N-n_s)}$ are the distances between the n_s sample nodes and the remaining $N - n_s$ nodes, and $\mathbf{C} \in \mathbb{R}^{(N-n_s)\times(N-n_s)}$ are the distances within the unsampled nodes. This method approximates the eigenvectors and eigenvalues of $\mathbf{W}$ by choosing n_s samples distributed across the image from the dataset of size N ($n_s \ll N$) [2, 3]. In our context, an approximate adjacency matrix $\widehat{\mathbf{W}}_N^k$ is computed for each image at time k, and a fusion step is then performed [3]. The fusion step consists of capturing the unique information given by each approximated graph into one fused graph ($\mathbf{W}_F$) by maximizing the distance between nodes (i.e. choosing those pixels that preserve most of the information):

$$\mathbf{W}_F = min(\widehat{w}^k_{N_{ij}}), \text{with } k = [1, 2],$$

where $\widehat{w}^k_{N_{ij}}$ represents the weight of the node for each instance of time k. In this sense, the learning of this approach is data-driven (uses a few n_s samples to learn) and it will be restarted from scratch for each dataset. To detect the change we use the approximated eigenvectors and eigenvalues found by NE from $\mathbf{W}_F$, as features to represent the change in the images (for more detail see algorithm 1 in [3]). Being that, the number of eigenvectors is equal to the number of samples (n_s) taken from an instance of time k, we build the change map, by selecting the scaled eigenvector that maximizes the mutual information (MI) [16] of this eigenvector with a binarized prior signal. This prior signal comes from the normalized differences between pre-event (I_{bf}) and post-event (I_{af}) images, by using $\mathcal{T}$ the global threshold (Otsu) [3]:

$$P_b = \mathcal{T}\left(\frac{I_{bf} - I_{af}}{I_{bf} + I_{af}}\right) + \mathcal{T}\left(\frac{I_{af} - I_{bf}}{I_{bf} + I_{af}}\right) \quad (1)$$

3. BLUE-NOISE SAMPLING ON GRAPHS

We now describe briefly the problem of sampling on graphs and the concept of BN on graphs.

Consider a graph $G = (V, E, \mathbf{W})$, and let $f \in \mathbb{R}^{|V|}$ be a signal defined on G, that is $f : V \mapsto \mathbb{R}$ can be considered a scalar function of V, where the value of f at the vertex $i \in V$ is denoted by $f(i)$ or f_i. When the signal f defined on G is known to be smooth with respect to the graph G, *sampling on graphs* aims at the characterization of *sampling sets* $S \subset V$, with cardinality $|S| = m$, such that the signal f can be exactly (or accurately) reconstructed from its values at S, that is $f_S = (f_i : i \in S)^T \in \mathbb{R}^m$. There are many ways to find a suitable sampling set $S \subset V$ [12]. In this work, however, we will focus only on a family of sampling sets, referred to as *blue-noise sampling sets* , widely used in the context of digital half-toning [17], but recently extended to graphs by [11].

At a basic level, a BN set is a subset of vertices S of V, whose elements are arranged in such a way that they are as far apart as possible from each other in terms of geodesic distances on G. As demonstrated in [11], these sampling sets lead to an accurate reconstruction of signals whose energy is mostly concentrated at the lowest eigenvectors of the graph Laplacian of G. This fact can be exploited in CD because in the approach proposed in [3] we construct graphs in such a way that the land-cover changes are detectable at the lowest eigenvectors. We now elaborate on our approach to CD based on BN on graphs and graph learning.

4. THE PROPOSED METHOD

Even though the aforementioned CD model compares favorably with recent state-of-the-art methods, its performance depends significantly on the accuracy of the NE to approximate $\mathbf{W}$ [3]. So, as stated in the introduction, we will use BN to enhance the NE approximation of the adjacency matrix. To do so, we first partition the image to regions related to the change by solving the isoperimetric problem, which is "to find a boundary of minimum perimeter enclosing maximal area" and the prior in equation 1 encloses the area related to the tentative change and some outliers. In doing so, we use the graph isoperimetric partitioning algorithm [14]. Once we got the regions, we generate the nodes $\mathbf{r} \in \mathbb{R}^{n_r}$ (where n_r is the total number of regions) as the mean value of each region. As a second step, we infer structural relationships among the obtained regions using graph learning (GL) with a smoothness prior [9, 13]. GL problem is to learn a graph topology as learning the Laplacian matrix ($\mathbf{L}$), such that the signal variation on the resulting graph ($\mathcal{Q}(\mathbf{L})$), is small, that is "the signal takes similar values at neighboring vertices" [9]. The measure of smoothness of a signal x in a graph is given by the scalar value: $\mathcal{Q}(\mathbf{L}) = \frac{1}{2}\sum_{i,j} w_{ij}(x(i) - x(j))^2$, where

w_{ij} is the ijth entry of matrix $\mathbf{W}$. This, on the one hand, smooths out outlying nodes related to regions where there is no change, and on the other hand, accentuates regions where change takes place. In this work, we use the recent approach in [13] to learn such a graph, where the authors leverage the desired graph sparsity to reduce computation and automatically select the parameters of the model. As in k-NN, the number of neighbors K (edges per node) is an input, without performing grid search over two parameters α and β. Then, the minimization problem of [13] is:

$$\begin{aligned} \min_{W} \quad & \textstyle\sum_i \sum_j W_{ij} Z_{ij} - \\ & \alpha \textstyle\sum_i \log(\sum_j W_{ij}) + \frac{\beta}{2} \|W\|_F^2 + \frac{c}{2} \|W - W0\|_F^2 \\ \text{s.t.} \quad & W_{ij} = W_{ji} \geq 0, i \neq j, diag(W) = \mathbf{0}, \end{aligned} \tag{2}$$

where $\mathbf{Z}$ is a pairwise distances matrix, α is a log prior constant ($> \alpha \rightarrow >$ weights in $\mathbf{W}$), β is a $\|W\|_F^2$ prior constant ($> \beta \rightarrow$ less sparsity in $\mathbf{W}$). Since matrix $\mathbf{Z}$ must contain information from the pre-event and pos-event images, we use the magnitude of difference image $|I_{bf} - I_{af}|$, computed for the mean value of the regions given by the graph partitioning algorithm. As a final step, we run the BN algorithm in [11] on the learned graph, and the BN samples are transformed into the image domain by selecting as samples the centroid of the regions associated with each of the vertices of the graph. These samples are next entered into NE to detect changes in the scene as shown in Figure 1.

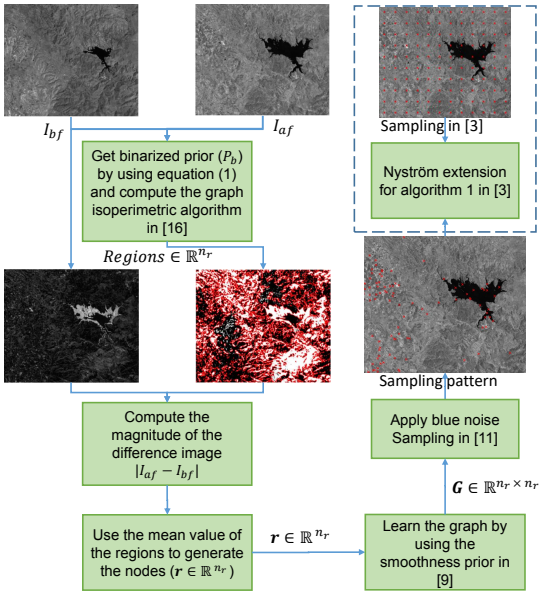

Fig. 1. Flow chart of the proposed method.

Figure 1 shows the flow chart of the proposed method. Note that the sampling pattern (red dots) in the image capture the structural information, as the lake and a city near the zone. However, some of the samples are close among them, and one may think that the BN is not providing well-separated nodes. Nonetheless, the reason for this proximity is that the graph and the image have different spaces ($Image \in \mathbb{R}^{m \times n}$ and $G : Image \mapsto \mathbb{R}^{N \times N}$, where $N = m * n$). In contrast, the uniformly spaced 2D grid sampling (dashed lines) does not capture information given by the structure of the image.

5. EXPERIMENTAL RESULTS AND DISCUSSION

5.1. Databases

We tested our approach in 14 real change detection scenarios captured by MS and SAR sensors presented in [3]. These datasets include events such as earthquakes, floods, wildfires, melted ice, farming, and building. In addition, these datasets include 4 multi-modal datasets (Toulouse, California, Bastrop, Gloucester-2) which combine SAR/MS images. For a more detailed description of the datasets please refer to [3].

5.2. Experimental set-up

We ran all the codes [1] in a server with 2 processors, Intel(R) Xeon(R) CPU E5-2650 v4 @2.20GHz, a total of 24 physical cores, 48 threads of processes, and 252 GB of RAM @2400 MHz. We compared the proposed method with our previous approach in [3]. We evaluated the change-detection map generated with respect to the ground truth by using the well-known Cohen's kappa coefficient (κ): $\kappa = \frac{p_o - p_e}{1 - p_e}$, where p_o is the observed agreement between predictions and labels (the overall accuracy), while p_e is the probability of random agreement.

The number of samples was set equal to the previous approach [3], the number of regions where set manually to cover the most of areas in the prior (equation (1)), and the number of neighbors (K) for the graph learning were set through exhaustive grid search. For the regions we use the graph analysis toolbox [2], for the graph learning we use the GSPBOX [3], and for the BN we use the implementation in [4].

The final results are tabulated in Table 1, where in overall the proposed method improves the results in 10 datasets out of 14. Bastrop and Mulargia datasets present a small percentage of improvement ($\approx 1\%$) because the priors used for these cases do not provide enough discrimination between change/no change regions. However, for the remaining 8 datasets proposed technique outperforms the competing state-of-the-art techniques rrR [4] and U-CD-HPT [8] presented in [3]. It is important to remark, that for the 4 multi-modal datasets the proposed methodology presents better results in 3 (Toulouse, Bastrop, and Gloucester-2) of them, which demonstrates the effectiveness of the sampling method based on graphs and the use of graph signal processing for graph learning. For the remaining 3 datasets (Alaska, Wenchuan, and California) the results are close to those reported in [3].

6. CONCLUSIONS

In this paper we introduced a graph-based sampling method for change detection, which builds on recent state-of-the-

[1] Available at: `github.com/DavidJimenezS`
[2] Available at: `leogrady.net/software/`
[3] Available at: `epfl-lts2.github.io/gspbox-html`
[4] Available at: `github.com/jhonygiraldo`

art graph signal processing tools, that is BN on graphs and GL. Our main contribution is a graph-based framework for sampling relevant nodes related to regions of remote sensing images. Our method models an image as regions, learn a graph by using smoothness prior over the regions, and apply BN to select relevant nodes. Experimental results showed that the proposed model outperformed recent state-of-the-art methods based on graphs (US-2D) [3], based on machine learning (U-CD-HPT) [8] , and based on probability (rrR) [4] in 10 datasets out of 14 in terms of the metric κ. According to the previous results and analysis, we conclude that the proposed algorithm for graph sampling in remote sensing images is a promising and robust method for change detection approaches based on graphs [3].

Acknowledgments This work was funded by the OMICAS program: "Optimización Multiescala In-silico de Cultivos Agrícolas Sostenibles (Infraestructura y validación en Arroz y Caña de Azúcar)", anchored at the Pontificia Universidad Javeriana in Cali and funded within the Colombian Scientific Ecosystem by The World Bank, the Colombian Ministry of Science, Technology and Innovation, the Colombian Ministry of Education, the Colombian Ministry of Industry and Tourism, and ICETEX, under grant ID: FP44842-217-2018 and OMICAS Award ID: 792-61187.

Table 1. Performance of the approaches in terms of $\boldsymbol{\kappa}$.

Dataset	**rrR** [4]	**U-CD-HPT** [8]	**US-2D** [3]	**BN**
Mulargia	0.7929	0.6984	0.9043	**0.9096**
Omodeo	**0.8112**	0.6414	0.2873	0.5903
Alaska	0.7368	0.8565	**0.8917**	0.8620
Madeirinha	0.6727	0.5048	0.8046	**0.8618**
Katios	−0.1790	0.2882	0.3196	**0.5285**
Atlantico	−0.0040	0.0970	0.4726	**0.5936**
San Franciso	0.1311	0.3143	0.4285	**0.6720**
Wenchuan	0.2380	−0.2730	**0.3239**	0.3158
Toulouse	0.1329	0.1200	0.1702	**0.2670**
Prince George	0.5322	0.4742	0.5442	**0.7284**
California	−0.1430	**0.3845**	0.3507	0.3418
Gloucester-1	0.1724	0.5794	0.2286	**0.6134**
Bastrop	0.000	0.8884	0.8875	**0.8887**
Gloucester-2	0.1693	0.0080	0.1562	**0.3583**

7. REFERENCES

[1] Mauro Dalla Mura, Saurabh Prasad, Fabio Pacifici, Paulo Gamba, Jocelyn Chanussot, and Jón Atli Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1585–1601, 2015.

[2] Geoffrey Iyer, Jocelyn Chanussot, and Andrea L Bertozzi, "A graph-based approach for data fusion and segmentation of multimodal images," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[3] David Alejandro Jimenez-Sierra, Hernán Darío Benítez-Restrepo, Hernán Darío Vargas-Cardona, and Jocelyn Chanussot, "Graph-based data fusion applied to: Change detection and biomass estimation in rice crops," *Remote Sensing*, vol. 12, no. 17, pp. 2683, 2020.

[4] Massimo Zanetti and Lorenzo Bruzzone, "A theoretical framework for change detection based on a compound multiclass statistical model of the difference image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 1129–1143, 2017.

[5] Mi Song, Yanfei Zhong, and Ailong Ma, "Change detection based on multi-feature clustering using differential evolution for landsat imagery," *Remote Sensing*, vol. 10, no. 10, pp. 1664, 2018.

[6] Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, and Brian Alan Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS journal of photogrammetry and remote sensing*, vol. 152, pp. 166–177, 2019.

[7] Redha Touati, Max Mignotte, and Mohamed Dahmane, "Multimodal change detection in remote sensing images using an unsupervised pixel pairwise based markov random field model," *IEEE Transactions on Image Processing*, 2019.

[8] L. T. Luppino, F. M. Bianchi, G. Moser, and S. N. Anfinsen, "Unsupervised image regression for heterogeneous change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9960–9975, 2019.

[9] Xiaowen Dong, Dorina Thanou, Michael Rabbat, and Pascal Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 44–63, 2019.

[10] Sanjiv Kumar, Mehryar Mohri, and Ameet Talwalkar, "Sampling methods for the nyström method," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 981–1006, 2012.

[11] Alejandro Parada-Mayorga, Daniel L Lau, Jhony H Giraldo, and Gonzalo R Arce, "Blue-noise sampling on graphs," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 3, pp. 554–569, 2019.

[12] Yuichi Tanaka, Yonina C Eldar, Antonio Ortega, and Gene Cheung, "Sampling signals on graphs: From theory to applications," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 14–30, 2020.

[13] Vassilis Kalofolias and Nathanal Perraudin, "Large scale graph learning from smooth signals," in *International Conference on Learning Representations*, 2019.

[14] Leo John Grady and Eric L Schwartz, *Isoperimetric graph partitioning for data clustering and image segmentation*, Boston University, Center for Adaptive Systems and Department of Cognitive , 2003.

[15] Charless Fowlkes, Serge Belongie, Fan Chung, and Jitendra Malik, "Spectral grouping using the nystrom method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 214–225, 2004.

[16] Thomas M Cover and Joy A Thomas, *Elements of information theory*, John Wiley & Sons, 2012.

[17] Daniel L Lau and Gonzalo R Arce, *Modern digital halftoning*, vol. 1, CRC Press, 2018.