Robust Clustering for Satellite Images Time-Series

Keywords : machine learning, statistics, geoscience, earth observation

1. THESIS PROJECT TEAM

a. Ph.D. Student

The candidate must own a master/engineering degree with knowledge in statistics and machine learning (having followed courses and/or an internship on these topics is strongly advised). Good coding skills in either Python or Matlab will also be required.

b. Supervision

The thesis will be held at SONDRA, CentraleSupelec. The supervision team is the following

Name	Institution	Role in the Thesis
Jean-Philippe OVARLEZ	SONDRA and ONERA	Director
Guillaume GINOLHAC	University of Savoie Mont-Blanc	Co-director
Chengfang REN	SONDRA, CentraleSupélec	Co-supervisor
Arnaud BRELOY	University of Paris X	Co-supervisor

2. CONTEXT AND MOTIVATION

Remote sensing data from Synthetic Aperture Radar (SAR) sensors offers a unique opportunity to record, to analyze, and to predict the evolution of the Earth. In the last decade, numerous satellite remote sensing missions have been launched (Sentinel-1, UAVSAR, TerraSAR X, etc.). This resulted in a dramatic improvement in the Earth image acquisition capability and accessibility. The growing number of observation systems allows now to build high temporal/spatial-resolution Earth surface images data-sets. This new scenario significantly raises the interest in time-series processing to monitor changes occurring over large areas. On the other hand, developing new algorithms to process such a huge volume of data represents a current challenge. Notably, the modern trend of deep-learning approaches shows its limits since most of this data is not annotated and corrupted by a problematic speckle noise (inherent to SAR images). Thus, statistical learning methods, that are able to leverage physical prior knowledge, appear suited to the task and lead to good performance in practice.

The derivation of novel statistical data processing techniques within this framework is of huge interest to the scientific community for applications such as deforestation assessment, urbanization monitoring, or land activity surveillance. A steering issue concerns the clustering (i.e., automatically sort) of spatial zones in the SAR images, which drives many current research works. However, exploiting the full spatio-temporal diversity of times-series is still an opened question we aim to address in this Ph.D. thesis.

3. TECHNICAL PROGRAM

The objective of this thesis is to develop and study new clustering methods for satellite image time-series (ITS). The proposed approach hinges on the robust estimation framework and differential geometry for the design of relevant distance between stacks of images. Indeed, robust statistical processing appears well suited to SAR ITS. Notably, we have shown its use for change detection in [1]. Additionally, distances built from a differential geometry perspective achieve outstanding performance for clustering purposes [2], but this approach has not been brought to the robust framework yet.

a. Design metrics to cluster

Clustering techniques are usually based on a choice of metrics/distance between objects of interest [3], that has a decisive impact on the clustering accuracy. In the framework of robust statistics [4], our aim is to use the information theory perspective to build meaningful distances on feature manifolds (such as local covariance matrices and textures). Notably, we aim to extend our work [5] for structured covariance manifolds (e.g. low-rank and/or Kronecker structured), as these parameters can efficiently reflect the underlying physics of SAR data. From a computational point of view, an interesting prospect will also be to consider the use of metric learning techniques [6] for manifolds that are too complex to handle theoretically.

b. Clustering in image time-series

Using the aforementioned metrics/distances this part aims to develop new clustering algorithms. Two approaches will be considered, depending on which dimension is to be clustered in the image stack:

- Temporal clustering with respect to space: cluster the pixels according to the patterns of evolution (impulsive, periodic or slow changes). In this case, the focus is not on the different objects present in the scene, but rather in their temporal behavior. The output of this process can be used for Change Detection (CD) purposes [1].
- Spatial clustering with respect to time: cluster the objects present in the scene using spatial and temporal data. In this case, the objective is twofold: distinguishing zones corresponding to different objects while taking into account their temporal evolution [7]. Such approach will be possible by appropriately integrating the structure of the image stack in the distances.

The methodologies developed will be assessed on real image time-series obtained from Sentinel-1, UAVSAR and TerraSAR-X missions [8][9].

References

[1] A. Mian, G. Ginolhac, J. Ovarlez et A. Atto, «New robust statistics for change detection in time series of multivariate SAR images,» *IEEE Transactions on Signal Processing,* 2018.

[2] S Said, H Hajri, L Bombrun, BC Vemuri, “Gaussian distributions on Riemannian symmetric spaces: statistical learning with structured covariance matrices”, IEEE Transactions on Information Theory 64 (2), 752-772

[3] A. Singh, A. Yadav et A. Rana, «K-means with Three different Distance Metrics,» *International Journal of Computer Applications,* vol. 67, n° %110, 2013.

[4] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, Complex Elliptically Symmetric Distributions: Survey, New Results and Applications, IEEE Transactions on Signal Processing, vol.60, issue.11, pp.5597-5625, 2012.

[5] A. Breloy, G. Ginolhac, A. Renaux, F. Bouchard, “Intrinsic Cramer-Rao Bounds for Scatter and Shape Matrices Estimation in CES Distributions”, IEEE Sig. Proc. Letters, vol. 26(2), pp 262-266, 2019.

[6] M. Harandi, M. Salzmann, R. Hartley, “Joint Dimensionality Reduction and Metric Learning: A Geometric Take”, Proceedings of the 34th International Conference on Machine Learning, PMLR 70:1404-1413, 2017.

[7] Alice Le Brigant, “Computing distances and geodesics between manifold-valued curves in the SRV framework”, arXiv:1601.02358v4

[8] UAVSAR https://uavsar.jpl.nasa.gov/

[9] SDMS https://www.sdms.afrl.af.mil/index.php?collection=ccd_challenge