Sujet de stage de fin d'étude R&D 2024

Sujet : Recherche efficace **et sans apprentissage** de motifs spatio-temporels dans des grands cubes de données satellites
Discipline : Informatique, Traitement du Signal et des Images
Mots-clefs : Traitement d'image, télédétection, séries temporelles, hiérarchies morphologiques, algorithmes efficaces

Contact

IRISA - site de l'UBS - Equipe OBELIX
François Merciol
francois.merciol@irisa.fr
IRISA, Bâtiment ENSIBS, Campus de Tohannic, 56000 Vannes

Environnement du stage

L'IRISA (Institut de Recherche en Informatique et Systèmes Aléatoires, `www.irisa.fr`), est une unité mixte de recherche (UMR 6074) en informatique, en traitement du signal et des images et en robotique. Sur la conjonction de ces thématiques, l'IRISA se positionne comme le grand laboratoire de recherche en Bretagne avec une présence affirmée sur les campus de Rennes (35), Vannes (56), et Lannion (22). Elle s'organise autour de 7 départements, dont celui "Signal, Image, Langage" dans lequel s'inscrivent les travaux de l'équipe OBELIX.

L'équipe OBELIX (Observation de l'environnement par imagerie complexe) mène ses recherches dans le domaine de l'intelligence artificielle (apprentissage automatique et vision par ordinateur) appliquée à l'observation de la Terre. Elle conçoit ainsi des solutions informatiques originales, efficientes et robustes pour faire face à la complexité inhérente aux données environnementales. En particulier, l'équipe a développé ces dernières années une expertise méthodologique dans les domaines de l'apprentissage profond, du transport optimal, et des hiérarchies morphologiques.

Cette équipe est principalement localisée à Vannes (56), et dispose de moyens de calcul (clusters CPU et GPU) qui pourront être mobilisés pendant le stage. Elle porte également le parcours de Master 2 GeoData Science du Master Erasmus Mundus "Copernicus Master in Digital Earth".

L'encadrement sera assuré par François Merciol (IRISA - UBS), avec l'aide de Sébastien Lefèvre (IRISA - UBS) et d'Antoine Masse (CLS).

Contexte applicatif et industriel

Depuis 2011, le Copernicus Land Monitoring Service (CLMS `https://land.copernicus.eu/`) fournit des produits pour la surveillance de l'état, des changements et des caractéristiques de la couverture/utilisation des terres végétalisées, non végétalisées, des variables biophysiques, des conditions de l'eau et de la cryosphère. Cette cartographie à large échelle utilisent les données d'observation à haute résolution de la terre (10 m de résolution spatiale, une acquisition toutes les semaines) gratuites et libres d'accès telles que Sentinel (1 et 2), ainsi que des données commerciales à très haute résolution spatiale (1 m de résolution spatiale, tous les ans).

Dans ce contexte, il est important de développer des algorithmes, des méthodes et procédés semi-automatiques (voire automatiques) afin de limiter au strict nécessaire le recours à des traitements humains au regard des masses de données manipulées. Des interventions humaines, même de courte durée, répétées à l'échelle de l'Europe voire du globe ont un effet majeur sur la capacité à fournit les produits Copernicus en un temps raisonnable. De plus, il est nécessaire d'optimiser les traitements informatiques pour réduire leur coût financier et énergétique. Dans ce contexte, l'unité de R&D du pôle Terre et Eau du groupe CLS cherche à concevoir des solutions (semi-)automatiques efficaces pour analyser de grands volumes de données d'observation de la Terre.

La recherche d'automatisation a donné lieu à une collaboration entre CLS et OBELIX depuis plusieurs années, afin de concevoir et déployer des solutions efficientes de cartographie automatisée large-échelle.

Contexte scientifique

Ainsi, dans le cadre d'une demande de l'Agence Européenne de l'Environnement, CLS et OBELIX ont conçu et déployé une chaîne de production originale de la cartographie des trames vertes à l'échelle continentale pour le compte du programme Copernicus. Pour faire face au volume de données à traiter (38 000 images, soit 120 To), et à la diversité des scènes étudiées, la solution développée s'est appuyée sur des algorithmes efficaces de caractérisation multi-échelle des pixels (profils d'attributs) à l'aide des hiérarchies morphologiques, et de classification semi-supervisée par une approche ensembliste de forêts aléatoires [3]. Elle a été implantée à l'aide de composants logiciels C++ diffusés sous licence libre : TRISKELE [1] et Broceliande [2]. Une attention particulière est portée à l'optimisation systématique de toutes les étapes du processus, y compris l'extraction des descripteurs [4]. Cette étape, centrale dans le processus de cartographie automatique, est souvent mise en oeuvre à l'aide des profils d'attributs calculés efficacement à l'aide des hiérarchies morphologiques [6].

Dans un autre contexte, l'équipe OBELIX a collaboré avec le CNES dans le cadre d'une étude R&T pour développer une solution efficace de recherche automatique par l'exemple dans des bases d'images satellites. Pour cela, elle a exploité les hiérarchies morphologiques pour calculer des histogrammes de formes (ou Pattern Spectra) qui permettent de mettre en oeuvre des algorithmes efficaces de recherche par l'exemple. La solution ainsi développée rend possible la recherche de motifs spatiaux de taille variable (et non connue a priori) dans une base de très grandes images [5]. Contrairement aux approches populaires en vision par ordinateur basées sur l'apprentissage automatique ou profond, elle ne s'appuie pas sur un entraînement préalable d'un modèle prédictif, et fonctionne sans recourir à des données annotées. Ce travail a abouti au démonstrateur Korrigan (Fig. 1).

Au vu de la pertinence des hiérarchies morphologiques et des outils qui en découlent (profils d'attributs, histogrammes de formes) pour élaborer des solutions efficaces d'analyse semi-automatique d'images satellites, leur extension aux séries temporelles d'images satellites a également été étudiée, au travers d'une thèse de doctorat conduite conjointement par l'équipe OBELIX, le CNES, et CLS [7].

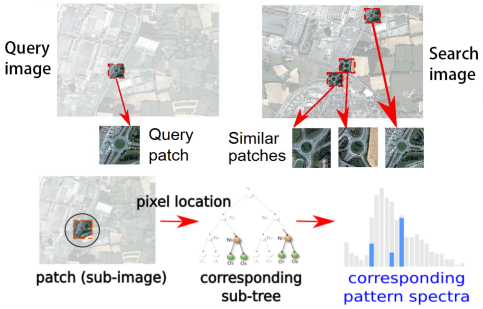

FIGURE 1 – Illustration d'une recherche par l'exemple avec KORRIGAN, une approche déterministe efficace développée par l'équipe OBELIX et le CNES [5]. Un opérateur sélectionne une zone d'intérêt dans une image satellite (*query image*) et souhaite obtenir des résultats semblables dans une image cible (*search image*). Pour ce faire, le modèle de recherche (*patch*) produit une représentation hiérarchique de laquelle sont extraits des sous-arbres (*corresponding sub-tree*). Ils produiront un sous-histogramme (*corresponding pattern spectra*) servant de signature à la recherche.

Sujet

En s'appuyant sur les développements précédents de l'équipe sur les hiérarchies morphologiques appliquées à l'analyse d'images et de séries temporelles d'images satellites, l'objectif du stage sera de définir une solution originale et efficace pour la recherche de motifs spatio-temporelles. **La solution à mettre en oeuvre ne s'appuiera pas sur l'apprentissage profond.** L'intérêt d'un tel choix, à contre-courant de l'engouement actuel pour l'IA, réside dans la capacité du système de recherche à fonctionner sans données d'entraînement (annotées ou non), sa faible empreinte carbone du fait de l'optimisation des ressources (calcul et mémoire) permise par l'utilisation de structures de données adaptées (arbres) et d'algorithmes efficaces, ces derniers permettant en outre le déploiement d'une solution interactive.

Plus précisément, alors que nos travaux initiaux conduits avec le CNES ont donné lieu à un démonstrateur où la recherche portait sur des motifs spatiaux, il s'agira ici d'ajouter une contrainte temporelle, ce qui permettra la recherche d'objets et/ou d'évènements représentés par des motifs spatio-temporels. Bien que des travaux préliminaires aient été menés pour fournir une définition de l'histogramme de forme pour des données spatio-temporelles [9, 8], les algorithmes n'ont pas été optimisés et la mise en oeuvre n'est pas efficace. Ce volet reste cependant primordial pour envisager un déploiement opérationnel sur des grands cubes de données (plusieurs

To) avec la société CLS (comme lors des contributions au programme Copernicus décrits précédemment). Dans ce contexte, la méthodologie sera validée sur plusieurs scénarios d'utilisation avec l'aide du partenaire CLS qui fournira les jeux de données.

Au cours du stage, le travail consistera à :
— Étude bibliographique des *pattern spectra* et de leurs applications
— Étude bibliographique de la recherche de motifs dans les séries temporelles d'images satellites
— Prise en main de l'environnement logiciel
— Proposition d'une méthodologie
— Prototypage de la solution
— Validation sur des jeux de données fournis par CLS

Profile recherché

Nous recherchons des candidats, en voie d'achèvement de leur cursus de Master ou école d'ingénieurs, qui affichent d'excellents résultats académiques, et les compétences suivantes :
— Algorithmique avancée et programmation (Python et/ou C++)
— Traitement d'image
— Notions ou intérêt pour la télédétection
— Bonnes compétences verbales et écrites en anglais scientifique
— Curiosité, rigueur et esprit d'équipe

Dates

Période de stage prévue durant le premier semestre 2024.

Débouchés

En fonction du niveau et de la motivation du ou de la stagiaire, une poursuite en thèse CIFRE avec la société CLS est envisagée pour septembre 2024. La thèse élargira les approches méthodologiques envisagées à l'apprentissage profond, dans le but de comparer les avantages et inconvénients des deux paradigmes, et d'élaborer une stratégie hybride le cas échéant.

Gratification

Le stage est rémunéré conformément à la législation.

Références

[1] François Merciol and al. Tree representations of images for scalable knowledge extraction and learning for earth observation. 2017. `https://gitlab.inria.fr/obelix/triskele/`.

[2] François Merciol and al. Broceliande is a tools for classification base on triskele and random forest. 2018. `https://gitlab.inria.fr/obelix/broceliande/`.

[3] François Merciol, Loïc Faucqueur, Bharath Bhushan Damodaran, Pierre-Yves Rémy, Baudouin Desclée, Fabrice Dazin, Sébastien Lefèvre, Antoine Masse, and Christophe Sannier. Geobia at the terapixel scale : Toward efficient mapping of small woody features from heterogeneous vhr scenes. *ISPRS International Journal of Geo-Information*, 8(1) :46, 2019.

[4] François Merciol, Minh-Tan Pham, Deise Santana, Antoine Masse, and Christophe Sannier. BROCELIANDE : a comparative study of attribute profiles and featureprofiles from different attributes. In *ISPRS*, Nice, France, June 2020.

[5] Behzad Mirmahboub, Jérôme Moré, David Youssefi, Alain Giros, François Merciol, and Sébastien Lefèvre. Fast Pattern Spectra using Tree Representation of the Image for Patch Retrieval. In *Discrete Geometry and Mathematical Morphology 2021*, pages 107–119, May 2021.

[6] Deise Santana Maia, Minh-Tan Pham, Erchan Aptoula, Florent Guiotte, and Sébastien Lefèvre. Classification of remote sensing data with morphological attributes profiles : a decade of advances. *IEEE geoscience and remote sensing magazine*, 9(3) :43–71, 2021.

[7] Caglayan Tuna. *Morphological Hierarchies for Satellite Image Time Series*. PhD thesis, Université Bretagne Sud, 2020.

[8] Caglayan Tuna, Alain Giros, François Merciol, and Sébastien Lefèvre. On Morphological Hierarchies for Image Sequences. In *ICPR 2020 – 25th International Conference on Pattern Recognition*, pages 1–8, Milan (virtual), Italy, January 2021.

[9] Caglayan Tuna, François Merciol, and Sébastien Lefèvre. Spatio-temporal object stability for monitoring evolving areas in satellite image time series. In *ISPRS (Virtual Conference)*, Nice, France, June 2020.