

Sujet de thèse CIFRE : Interprétation automatique de données géophysiques par techniques d'apprentissage

Contexte

Le géoradar ou Ground Penetrating Radar (GPR) est une technologie permettant de sonder les sols à la recherche d'objets enfouis ou étudier la composition du sol (différentes couches constituantes et leurs proportions par exemple). Il s'agit d'un système radar émettant une onde électromagnétique pénétrant le sol et se réfléchissant sur les différents éléments le constituant. En captant les signaux réfléchis, on peut obtenir un signal appelé radargramme et qui est analysé pour étudier le sol en question. L'amélioration récente des technologies en termes d'antennes ont permis une réduction significative de la taille des géoradars. Ainsi il est envisageable de considérer un scénario de géoradar aéroporté (monté sur un drone) afin d'analyser les sols ainsi que les pans de montagnes de manière plus extensive.

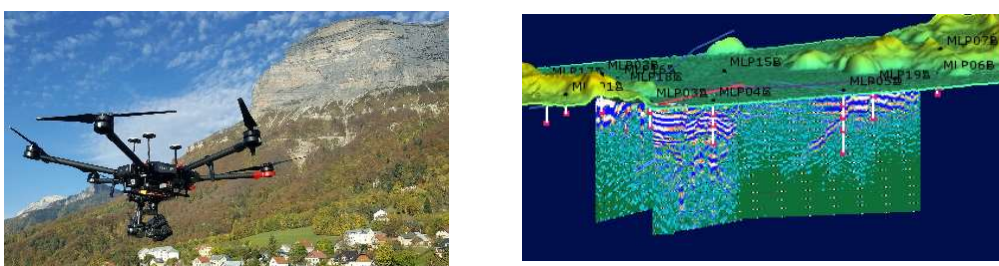


Figure 1 Scénario du GPR aéroporté

Dans ce cadre, le projet s'intéresse à l'étude radargrammes obtenus en milieu montagneux dans le cadre de mission de protection contre les risques naturels gravitaires. Cette information est capitale pour prévoir la chute de rochers ainsi que des glissements de terrain dont la fréquence augmente avec le réchauffement climatique et éviter des dégâts potentiels aux infrastructures de montagne telles que les routes, ponts, bâtiments et autres infrastructures liés aux activités économiques telles que le tourisme. Le scénario du géoradar aéroporté étant éloigné du scénario classique du géoradar plaqué au sol, il est ainsi nécessaire de prendre en compte les spécificités de ce nouveau mode d'acquisition. Une transposition directe des outils de traitements développés pour le géoradar plaqué au sol n'est donc pas possible.

Le but général dans ce contexte est de détecter / classifier les différentes structures géologiques ou objets présents dans le sous-sol. Pour réaliser cette opération, Géolithe a l'intention d'utiliser des techniques d'intelligence artificielle (IA) nécessitant des données préalablement labélisées. Ce travail de labellisation a été initié au sein de l'entreprise et la base de données est mise à jour continuellement.

Problématiques :

Malheureusement, il est très difficile d'appliquer directement les techniques d'IA à cause de plusieurs paramètres :

- Les images provenant des données GPR sont déjà dans le cas classique assez bruitées (voir Fig. 2) entraînant un rapport signal à bruit peu important. Ce problème est encore plus prononcé pour les géoradar aéroporté car ils sont situés plus loin de la surface. De plus, ces images GPR aéroportés comportent un certain nombre de données aberrantes qui peuvent réduire fortement la performance des algorithmes d'IA.
- La diversité des données est assez pauvre. En effet, une seule gamme de fréquence est utilisée et il n'y a pas d'information polarimétrique ce qui rend la distinction entre les objets délicate.
- Les données sont labélisées par des experts mais pour un non spécialiste, il n'est pas toujours évident de distinguer les différences et il est donc probable que la phase d'apprentissage soit cruciale pour bien distinguer les différentes réponses du sous-sol.
- Même si les données labélisées par géolithe sont importantes, elles seront sûrement insuffisantes pour entraîner un algorithme d'IA n'ayant pas été pré-entraîné.

Pour utiliser efficacement les techniques d'IA, il va falloir bien prendre en compte ces différentes problématiques et faire évoluer les algorithmes existants pour qu'ils s'adaptent à la caractéristique des données GPR. Plus particulièrement, il sera difficile d'utiliser les données brutes directement.

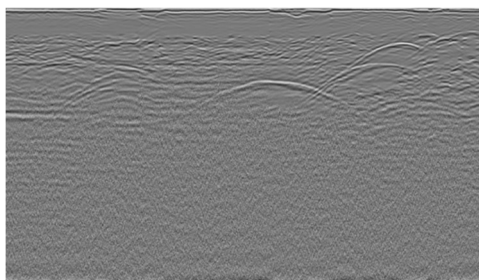


Figure 2 Exemple de radargramme brut

Objectifs de la thèse

Les objectifs de la thèse sont doubles :

- Dans un premier temps il s'agit de trouver un espace de représentation des données (features) permettant de mieux faire ressortir les différences entre les différentes classes des données labélisées. On s'intéressera notamment à des représentations par matrices de covariances qui sont une solution apportant des bonnes performances dans des applications liées au radar (Mian et al., 2019)
- Dans un second temps, l'objectif est de développer des algorithmes efficaces et adaptés à cet espace de représentation. Notamment, les features obtenues peuvent vivre un espace de représentation non-euclidien (comme c'est le cas pour les matrices de covariances (Arsigny et al., 2006)) et il sera nécessaire de prendre en compte cet aspect à l'aide d'outils tels que la géométrie riemannienne.
- Considérer d'un point de vue théorique les réseaux ainsi obtenus et apporter de la robustesse dans les architectures utilisées face aux contraintes évoquées en imagerie GPR.

Plan de travail :

- Étude de l'état de l'art en apprentissage en vision par ordinateur (Guo et al., 2016) et GPR (Daniels, 2004) afin de proposer une première solution basée sur des architectures traditionnelles et étude des limites dans le cadre du GPR
- Trouver une représentation permettant de mettre en évidence la différence entre bruit et signaux à classifier. Pour ce faire, nous nous intéresserons dans un premier temps à augmenter les données d'images à l'aide de filtres convolutifs obtenus en utilisant les premières couches de réseaux pré-entraînés dans des problèmes de classification d'images (MobileNet (Sandler et al., 2018), AlexNet (Krizhevsky et al., 2012), etc). Dans un second temps, on cherchera à obtenir une information plus concise à l'aide de descripteurs statistiques que l'on calculera sur les données augmentées. Plusieurs choix de descripteurs seront explorés (matrices de covariance ou sous-espaces entre autres).
- Utilisation de méthodes de la littérature pour classifier les features obtenues en prenant en compte leur caractère non-euclidien avec l'aide de la géométrie riemannienne. Dans ce cadre nous nous intéresserons à l'utilisation de classifieurs basés sur la notion de distance (MDM, K-NN), de classifieurs sur espace tangent (Barachant et al., 2013) et sur des architectures de réseaux tels que SPDnet (Huang & Gool, n.d.) et GrassmanNet (Huang et al., 2018).
- Afin d'apporter de la robustesse à ces méthodes nous considérerons la modification de ces réseaux de neurones à des modèles de géométrie riemannienne statistiques et rang-faibles (Bouchard et al., n.d.) qui sont connus pour être plus robustes dans le cas de données très bruitées tel qu'en GPR. Il sera nécessaire pour ce faire d'obtenir de nouveaux types de couches de réseaux et de dériver à nouveau l'expression du gradient par back-propagation de ceux-ci.
- Validation sur des données expérimentales obtenues sur des scénarios réels

Il sera possible de tester les méthodologies développées dans le cadre d'autre données géophysiques telles que des données issues de capteurs sismiques passifs pour détecter des événements (chute de pierre, tremblement de terre, etc)

Environnement scientifique

La thèse se déroulera au sein de deux organisations : Géolithe et le LISTIC :

- Géolithe est un Bureau d'Ingénieurs conseils en géologie, géophysique, géotechnique et génie civil, spécialisé dans les domaines des risques naturels, des infrastructures et aménagements, de l'environnement, des gisements et ressources.
- Le Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance (LISTIC) est une Unité de Recherche (UR) de l'Université Savoie Mont Blanc (USMB) dont les compétences se concentrent autour de l'apprentissage automatique, la spécification et le développement de théories, d'algorithmes et de systèmes pour extraire et gérer les connaissances, notamment dans les domaines de l'observation de la Terre

Dans le cadre du projet, Géolithe fournira les données d'apprentissage et de validation (profils radar en montagne), nécessaires au projet, ainsi que l'expertise spécifique à l'interprétation des données géoradar. Le LISTIC apportera une expertise sur l'apprentissage automatique et sur le développement de nouveaux algorithmes non-conventionnels (et plus robustes) afin de traiter les données compliquées en GPR.

Lieux :

- Laboratoire LISTIC, Annecy
- Géolithe, Grenoble

Encadrement :

- Ammar Mian, MCF Univ. Savoie Mont-Blanc
- Guillaume Ginolhac, PR Univ. Savoie Mont-Blanc
- Nickolas Stelzenmuller, Ingénieur de Recherche Géolithe

Profil du candidat

- Master recherche ou ingénieur avec une expérience en apprentissage statistique
- Intérêt pour les mathématiques appliquées (statistiques, algèbre linéaire)
- Compétences : vision par ordinateur, statistiques, apprentissage automatique, notamment deep learning
- Programmation : Python, frameworks de Deep Learning (Tensorflow, PyTorch)

Procédure de candidature

Envoyer un mail à :

ammar.mian@univ-smb.fr, guillaume.ginolhac@univ-smb.fr et nickolas.stelzenmuller@geolithe.com

Références

Arsigny, V., Fillard, P., Pennec, X., & Ayache, N. (2006). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1), 328–347. <https://doi.org/10.1137/050637996>

Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2013). Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing*, 112, 172–178. <https://doi.org/10.1016/j.neucom.2012.12.039>

Bouchard, F., Breloy, A., Ginolhac, G., Member, S., Renaux, A., & Pascal, F. (n.d.). *A Riemannian Framework for Low-Rank Structured Elliptical Models*.

Daniels, D. J. (2004). *Ground Penetrating Radar-2nd Edition Edited by*. <http://www.mathcad.com>.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, 187, 27–48. <https://doi.org/10.1016/j.neucom.2015.09.116>

Huang, Z., & Gool, L. van. (n.d.). *A Riemannian Network for SPD Matrix Learning*. www.aaai.org

Huang, Z., Wu, J., & van Gool, L. (2018). Building deep networks on grassmann manifolds. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, 1, 3279–3286.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25, 1097-1115-undefined. <http://code.google.com/p/cuda-convnet/>

Mian, A., Ginolhac, G., Ovarlez, J.-P., & Atto, A. M. (2019). New Robust Statistics for Change Detection in Time Series of Multivariate SAR Images. *Transactions on Signal Processing*, 67(2), 520–534. <http://ieeexplore.ieee.org>.

