



© IGN – Scan LiDAR par cartographie mobile terrestre



© IGN – Scan LiDAR HD par cartographie mobile aérienne

Offre de thèse

Reconnaissance de lieu à grande échelle à partir de nuages de points 3D

Application à la lutte contre la désinformation

Objet de la thèse

Le projet de thèse porte sur la reconnaissance de lieu à grande échelle à partir de nuages de points 3D (en anglais “Point Cloud Based Large-Scale Place Recognition”), avec pour application la géolocalisation de données image 3D. Sans connaissance d’une position initiale, géolocaliser un contenu image repose sur des approches d’indexation et de recherche par similarité de contenus dans un référentiel géolocalisé. Cette thèse propose d’étudier ce type d’approche en exploitant les cartographies 3D reposant sur des campagnes d’acquisition (notamment LiDAR), qui se démocratisent et dont la richesse en matière de géométrie et de sémantique les rend attractives, mais complexes à manipuler étant donné leur volume et diversité.

Contexte

Les domaines applicatifs de la reconnaissance de lieu à partir d’images sont nombreux, nous traiterons ici du cas de la géolocalisation de séquences vidéo amateur comme outil de certification pour la lutte contre la désinformation. Massivement diffusées sur les réseaux sociaux et sur le web, les vidéos amateur relayant une information ou un événement sont désormais très nombreuses, avec parmi elles des contenus qui relèvent de *fake news*, c’est-à-dire sortis de leur contexte d’origine pour exprimer une mauvaise ou une fausse information. Pour lutter contre cette forme de désinformation, plusieurs médias, tels que France TV, ont mis en place une cellule de *fact checking* d’images et de vidéos, qui analyse, vérifie et certifie ces flux d’images¹. Ce travail complexe est réalisé à la main et gagnerait à être automatisé en exploitant les outils de l’intelligence artificielle. Parmi les critères analysés, la vérification de la géolocalisation y a été reconnue comme essentielle pour savoir si le contenu a été proposé dans le bon contexte d’actualité pour expliquer au mieux ce qu’il se passe. C’est dans ce contexte collaboratif entre l’IGN et France TV que nous nous focalisons sur ce critère de géolocalisation avec la volonté d’exploiter les meilleurs référentiels de géoréférencement d’aujourd’hui pour proposer des solutions automatiques de géolocalisation à grande échelle, qui peuvent entre autres contribuer au *fact checking* de l’information visuelle.

Descriptif du sujet de thèse

Sans connaissance a priori d’une position initiale suffisamment précise, géolocaliser un contenu image, qu’il soit 2D, 2D+t, 3D ou nD, se traite par le biais de l’indexation et de la recherche par

¹ Cellule de fact checking de France TV : <https://www.francetvlab.fr/articles/les-revelateurs-de-francetv-la-cellule-de-fact-checking-video-et-images-de-france-televisions>

contenu à grande échelle : on commence par rechercher un ou des lieux similaires en termes de similarité de contenu dans un ensemble de même nature et potentiellement grand, qui cartographie l'environnement et où les données sont géolocalisées. Les données géolocalisées similaires retrouvées fournissent une ou plusieurs localisations candidates, qu'il est ensuite possible, si nécessaire, de raffiner, par exemple par estimation de la pose à l'aide d'une étape de recalage.

Avec la prolifération des bases d'apprentissage d'images de lieux (ex. Google Landmarks) et l'engouement pour l'automatisation des systèmes mobiles (navigation autonome, drones, etc.), la littérature sur la localisation basée image, et ses multiples déclinaisons ("landmark retrieval", "place recognition", etc.), s'est beaucoup développée [Pion et al, 2020], avec des objectifs différents en matière de précision de la localisation [Blettery et al, 2021] et de source d'information exploitée [Piasco et al, 2021]. La plupart des approches manipule des images 2D, pour lesquelles l'extraction de caractéristiques visuelles est un processus bien connu depuis des dizaines d'années. L'exploitation d'une information tridimensionnelle décrivant la scène est moins répandue, partant souvent de l'hypothèse qu'avec l'information 3D (LiDAR, SLAM, photogrammétrie, etc.) une localisation initiale est disponible (ex. GPS), ce qui ramène la localisation à un problème de recalage ("registration") de données 2D ou 3D dans un nuage 3D [Uy et al, 2018] [Huang et al, 2021].

Dans cette thèse, nous avons choisi d'exploiter cette information 3D pour répondre au problème de la géolocalisation de contenus, partant du principe que 1/ les sources de données 3D cartographiant l'environnement sont de plus en plus répandues, de plus en plus précises et de natures variées (différents LiDAR, SfM, photogrammétrie, capteurs RGB-D, etc.) et 2/ ces données géométriques fournissent des représentations robustes aux changements de la scène telles que le changement de point de vue ou l'écart temporel (nuit/jour, saison, etc.), qui continuent à pénaliser les approches reposant sur la photométrie des images 2D [Uy et al, 2018]. Lorsqu'il s'agit de rechercher dans un grand volume de référence 3D sans connaissance d'une position initiale, les techniques de l'état de l'art sont généralement rassemblées sous la dénomination "Point Cloud Based Large-Scale Place Recognition", qui reposent sur la reconnaissance de lieu à grande échelle par le biais de la recherche par similarité dans un grand volume de nuages 3D de référence, à partir d'un nuage 3D en requête.

Quelles que soient les données manipulées, la reconnaissance de lieu par similarité à grande échelle requiert la description et la structuration des données de la requête et de base de référence. Ces deux problèmes sont traités dans la thèse, et nous les décomposons selon trois axes de recherche :

1. **Description mono-source.** Dans un premier temps, le problème de la recherche de lieu sera traité en considérant une seule source de données 3D, des nuages LiDAR d'un type donné, pour étudier voire améliorer la robustesse des meilleures approches face aux données considérées et au volume. Nous étudierons comment décrire ces nuages avec pour objectif la recherche par similarité dans la base de référence. L'état de l'art sur la description de nuages de points est déjà vaste, avec des solutions faites main et apprises, et réparties dans le domaine de l'indexation et de la recherche d'objets 3D [Bold et al, 2019] et du recalage de nuages de points 3D [Zhou et al, 2020][Huang et al, 2021]. Lorsqu'il s'agit de décrire pour rechercher dans un grand volume 3D, les techniques les plus récentes s'appuient sur l'apprentissage profond de descripteurs locaux et globaux extraits dans l'ensemble de la scène [Liu et al, 2019][Hui et al, 2021][Xia et al, 2021][Komorowski, 2021][Hui et al, 2022] et font suite à la 1ère proposition sur le sujet, PointNetVLAD [Uy et al, 2018].
2. **Description multi-source.** Si l'on considère en requête une séquence d'images provenant d'un appareil standard (smartphone classique), le nuage de points 3D correspondant pourra être obtenu par Structure-from-Motion, il ne sera a priori pas de même nature que le nuage de référence. Par ailleurs, les sources de représentations 3D de l'environnement sont désormais assez nombreuses et variées : différents types de LiDAR (Riegl, Velodyne, aérien HD), SLAM, photogrammétrie, etc. Exploiter les modalités disponibles apporterait de la robustesse dans la recherche de lieu tout en étant exploitable à plus grande échelle. C'est un problème répandu en télédétection [Zhang et al, 2022], mais encore peu traité lorsqu'il s'agit de nuages de points 3D : les quelques approches considérant des nuages multi-source reposent massivement sur l'estimation de la transformation entre nuages par recalage [Huang et al, 2021], ce qui ne serait pas applicable ici. Plusieurs pistes de recherche pourront être considérées : traiter chaque modalité séparément et fusionner a posteriori les réponses candidates, ou plus ambitieux, être en mesure d'apprendre une description qui soit invariante à la modalité, dite inter-domaines, comme

c'est le cas avec certaines approches en télédétection [Zhang et al, 2022] ou en indexation de contenus patrimoniaux comme nous l'avons déjà étudié dans [Gominski et al, 2021].

- 3. Indexation et recherche à grande échelle.** Rechercher un lieu par similarité de descripteurs à grande échelle suppose de parcourir efficacement un grand volume de descripteurs multidimensionnels. Étonnamment, les meilleures approches actuelles de "Point Cloud Based Large-Scale Place Recognition" [Liu et al, 2019][Hui et al, 2021][Xia et al, 2021][Komorowski, 2021][Hui et al, 2022] sont évaluées sur une partie du dataset Oxford Robotcar [Maddern et al, 2016] qui correspond à une couverture LiDAR de quelques 10aines de km associée à approx. 100M de points pré-traités ; comparativement, les acquisitions en cartographie mobile terrestre faites sur Paris par l'IGN couvrent un linéaire de 200 km pour 70 milliards de points bruts. Si ses recherches se sont focalisées sur la richesse et discriminance des descripteurs proposés et non sur leur structuration pour une recherche la plus rapide possible, il semble indispensable de s'y intéresser pour passer à l'échelle, en termes de temps de recherche mais aussi de robustesse des descriptions face au volume. Les techniques d'indexation récentes les plus performantes reposent sur une recherche approximative par tables de hachage, déjà bien installées en indexation d'images [Rodrigues et al, 2020] et devenues populaires en télédétection [Reato et al, 2019]. En fonction des caractéristiques des descripteurs appris, il sera donc nécessaire de mettre en place une approche de représentation en code de hachage binaires, facilement distribuable et permettant une recherche efficace dans les descripteurs de nuages de points. Parmi les pistes à considérer, nous nous intéresserons aux approches hiérarchiques [Zhang et al, 2021-a], qui semblent fournir actuellement les meilleurs résultats.

Jeux de données considérés

Il est nécessaire de disposer de jeux de données à grande échelle avec une vérité terrain, d'une part pour valider qualitativement et quantitativement les approches proposées et d'autre part pour fournir une vérité terrain aux approches fondées sur des méthodes d'apprentissage. Les travaux reposeront sur les benchmarks publics internationaux mono-source et multi-source du domaine (ex. [Maddern et al, 2016] [Zhang et al, 2021-b]) mais aussi sur les jeux de données de l'IGN comme la cartographie 2D et 3D de l'intégralité de Paris par cartographie mobile terrestre² et la campagne LiDAR HD aérien en cours d'acquisition à l'échelle nationale³.

Profil du candidat

Bac+5 en informatique, mathématiques appliquées ou géomatique (master ou école d'ingénieur). Une bonne formation en apprentissage automatique est requise, avec des connaissances appréciées en vision par ordinateur 3D ou en indexation d'images. Le candidat retenu doit posséder de bonnes compétences en programmation (Python, C/C++).

Bien que la maîtrise du français ne soit pas requise, la maîtrise de l'anglais est nécessaire. La curiosité, l'ouverture d'esprit, la créativité, la persévérance et la capacité à travailler en équipe sont aussi des compétences personnelles clés recherchées.

Seuls les étudiants ressortissants de l'Union Européenne, du Royaume-Uni ou de la Suisse, sont éligibles à ce projet de thèse.

Organisation

Début : dernier trimestre 2022

Lieu : la thèse sera réalisée en région parisienne au laboratoire LaSTIG, situé à Saint-Mandé (73 avenue de Paris, métro Saint-Mandé, ligne 1) dans les locaux de l'IGN. Le(a) doctorant(e) sera rattaché(e) à l'Ecole Doctorale MSTIC (ED 532).

L'IGN (Institut national de l'information géographique et forestière) est un établissement public administratif dépendant du ministère de la transition écologique ; c'est l'opérateur national de

² Cartographie mobile terrestre Stéréopolis IGN : <https://www.ign.fr/institut/donnees-collectees-par-le-vehicule-stereopolis>

³ LiDAR HD IGN : <https://www.ign.fr/institut/lidar-hd-vers-une-nouvelle-cartographie-3d-du-territoire>

référence pour la cartographie du territoire français. Le LaSTIG⁴, Laboratoire en Sciences et Technologies de l'Information Géographique pour la ville intelligente et les territoires durables, est une unité mixte de recherche rattachée à l'Université Gustave Eiffel, à l'IGN et à l'École d'ingénieurs de la ville de Paris (EIVP). C'est une structure de recherche unique en France et même en Europe, regroupant approx. 80 chercheurs, qui couvrent l'ensemble du cycle de vie de la donnée géographique ou spatiale, de son acquisition à sa visualisation, en passant par sa modélisation, son intégration et son analyse ; parmi eux une trentaine de chercheurs exercent en analyse d'images, vision par ordinateur, apprentissage automatique, photogrammétrie et télédétection.

Les chercheurs et doctorants du LaSTIG peuvent être impliqués dans les activités d'enseignement de l'école d'ingénieurs de l'IGN, l'ENSG (École Nationale des Sciences Géographiques), qui offrent l'accès à des étudiants de premier et deuxième cycles d'excellente qualité dans les domaines liés aux sciences de l'information géographique : géodésie, photogrammétrie, vision par ordinateur, télédétection, analyse spatiale, cartographie, etc.

Comment candidater

Avant le 13 avril 2022, veuillez envoyer aux deux contacts en un seul fichier PDF les documents suivants :

- Un CV détaillé
- Une lettre de motivation centrée sur le sujet
- Notes et rangs sur les 3 dernières années d'étude
- Les coordonnées de 2 personnes référentes qui peuvent vous recommander

Contacts

- Laurent Caraffa – Laurent.Caraffa@ign.fr
Chargé de recherche au LaSTIG (encadrant de la thèse), IGN, Université Gustave Eiffel
- Valérie Gouet-Brunet – Valerie.Gouet@ign.fr
Directrice de recherche au LaSTIG (directrice de la thèse), IGN, Université Gustave Eiffel

Références

[Blettery et al, 2021] E. Blettery, N. Fernandes and V. Gouet-Brunet, How to spatialize geographical iconographic heritage, SUMAC'21 Workshop @ ACM Multimedia 2021, Oct 2021, Chengdu, China.

[Bold et al, 2019] N. Bold, C. Zhang and T. Akashi, 3D Point Cloud Retrieval With Bidirectional Feature Match, in *IEEE Access*, vol. 7, pp. 164194-164202, 2019.

[Gominski et al, 2021] Dimitri Gominski, Valérie Gouet-Brunet, Liming Chen, Connecting Images through Sources: Exploring Low-Data, Heterogeneous Instance Retrieval, *Remote Sensing (MDPI)*, 2021, 13 (16), pp.3080.

[Huang et al, 2021] Xiaoshui Huang, Guofeng Mei, Jian Zhang and Rana Abbas, A comprehensive survey on point cloud registration, *Computer Science ArXiv*, Published 3 March 2021.

[Hui et al, 2021] Le Hui, Hang Yang, Mingmei Cheng, Jin Xie, Jian Yang, Pyramid Point Cloud Transformer for Large-Scale Place Recognition, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 6098-6107.

[Hui et al, 2022] L. Hui, M. Cheng, J. Xie, J. Yang and M. -M. Cheng, Efficient 3D Point Cloud Feature Learning for Large-Scale Place Recognition, in *IEEE Transactions on Image Processing*, vol. 31, pp. 1258-1270, 2022.

[Khokhlova et al, 2022] Margarita Khokhlova, Nathalie Abadie, Valérie Gouet-Brunet and Liming Chen, GisGCN: A Visual Graph-Based Framework to Match Geographical Areas through Time, *ISPRS International Journal of Geo-Information, MDPI*, 11 (2), pp.97, 2022.

[Komorowski, 2021] J. Komorowski, MinkLoc3D: Point Cloud Based Large-Scale Place Recognition, in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2021 pp. 1789-1798.

[Liu et al, 2019] Zhe Liu, Shunbo Zhou, Chuanzhe Suo, Peng Yin, Wen Chen, Hesheng Wang, Haoang Li, Yunhui Liu, LPD-Net: 3D Point Cloud Learning for Large-Scale Place Recognition and Environment Analysis, in

⁴ Site web du LaSTIG : <https://www.umr-lastig.fr/>

- 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019 pp. 2831-2840.
- [Maddern et al, 2016] W. Maddern, G. Pascoe, C. Linegar and P. Newman, "1 Year, 1000km: The Oxford RobotCar Dataset", *The International Journal of Robotics Research (IJRR)*, 2016.
- [Piasco et al, 2021] Nathan Piasco, Désiré Sidibé, Valérie Gouet-Brunet, Cédric Demonceaux, Improving Image Description with Auxiliary Modality for Visual Localization in Challenging Conditions *International Journal of Computer Vision*, Springer Verlag, 129 (1), pp.185-202, 2021.
- [Pion et al, 2020] Noé Pion, Martin Humenberger, Gabriela Csurka, Yann Cabon, and Torsten Sattler, Benchmarking Image Retrieval for Visual Localization, In the International Conference on 3D Vision, 2020.
- [Reato et al, 2019] T. Reato, B. Demir and L. Bruzzone, An Unsupervised Multicode Hashing Method for Accurate and Scalable Remote Sensing Image Retrieval, in *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 2, pp. 276-280, Feb. 2019.
- [Rodrigues et al, 2020] Rodrigues, J., Cristo, M. & Colonna, J.G. Deep hashing for multi-label image retrieval: a survey. *Artif Intell Rev* 53, 5261–5307 (2020).
- [Uy et al, 2018] Mikaela Angelina Uy, Gim Hee Lee, PointNetVLAD: Deep Point Cloud Based Retrieval for Large-Scale Place Recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4470-4479.
- [Xia et al, 2021] Yan Xia, Yusheng Xu, Shuang Li, Rui Wang, Juan Du, Daniel Cremers, Uwe Stilla, SOE-Net: A Self-Attention and Orientation Encoding Network for Point Cloud Based Place Recognition, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 11348-11357.
- [Zhang et al, 2021-a] Y. Zhang, C. Peng, J. Zhang, X. Liu, S. Pu and C. Chen, Hierarchical Deep Hashing for Fast Large Scale Image Retrieval, *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 3837-3844, 2021.
- [Zhang et al, 2021-b] Jianing Zhang; Jinzhi Zhang; Shi Mao; Mengqi Ji; Guangyu Wang; Zequn Chen; Tian Zhang et al., "GigaMVS: A Benchmark for Ultra-large-scale Gigapixel-level 3D Reconstruction," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [Zhang et al, 2022] Zhang, X., Leng, C., Hong, Y., Pei, Z., Cheng, I. and Basu, A., Multimodal Remote Sensing Image Registration Methods and Advancements: A Survey, *Remote Sensing* 2021, 13, 5128.
- [Zhou et al, 2020] J Zhou, MJ Wang, WD Mao, ML Gong, and XP Liu. SiamesePointNet: A siamese point network architecture for learning 3d shape descriptor. In *Computer Graphics Forum*, volume 39, pages 309–321. Wiley Online Library, 2020.