Reconnaissance de lieux à grande échelle à partir de nuages de points

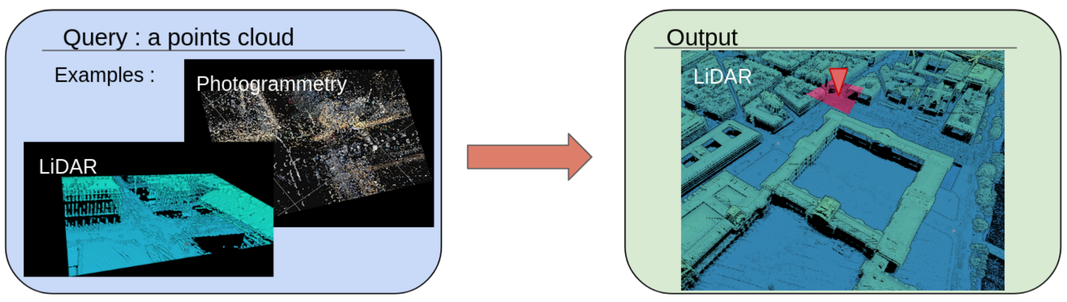
Application à la prévention de fake-news

Zede Chahine-Nicolas
Laboratoire LASTIG

Contexte

La plupart des approches de **reconnaissance de lieux basée image** s'appuient sur des bases de données de référence 2D.
Avec la démocratisation des données 3D, on souhaite exploiter l'indexation et la recherche dans une scène 3D pour la géolocaliser. Nous nous concentrons sur les modèles 3D reposant sur les nuages de points (LiDAR, photogrammétrie, etc.).

Objectifs



- **Géolocaliser un nuage de point**
- **S'adapter aux méthodes d'acquisition**
- **L'appliquer à grande échelle**
- **Évaluer la solution dans le contexte de la prévention de fake news de contenus vidéo (FranceTV, Gendarmerie Nationale)**

Stratégies

Pour retrouver la position d'un nuage en le comparant à une référence, nos axes de recherche se divisent en trois parties :

- **description mono-source**
  L'étude est d'abord menée avec une seule source de données (LiDAR) pour tester et évaluer les meilleures approches de description de nuages de points, comme PointNetVLAD [4] ou plus récemment LoGG3D-NET [5].
- **description multi-source**
  Ensuite, l'étude vise à s'étendre aux autres formats de production de données 3D (SLAM, photogrammétrie etc...) pour assurer des résultats robustes aux changements du type d'acquisition.
- **Indexation et recherche à grande-échelle**
  La recherche à grande échelle nécessite d'adapter les méthodes pour avoir des temps de recherche raisonnables (description compacte, structure d'index).

Notre méthode

Prenant inspiration dans la méthode de recherche par DSI (Derivative Search Index) [8], utilisée dans la recherche de corpus de textes, nous proposons une adaptation aux nuages de points. Ainsi, ici un nuage de points 3D en entrée donne une liste d'IDs des places les plus probables du nuage dans un dataset de référence, selon l'architecture ci-après.

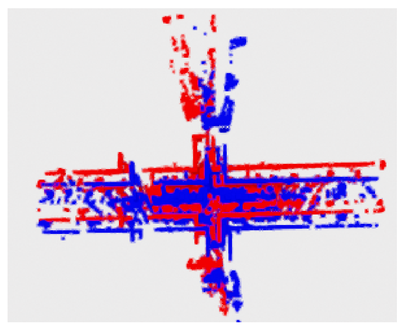
Architecture du modèle proposé

- **3D Encoder**, donne une représentation implicite du nuage de points. Cette partie est gelée lors de l'entraînement. Pour avoir un encodage adapté à la 3D, on utilise le SPT (Sparse Pyramid Transformers) de GD-MAE [7] qui construit un descripteur multi-échelle.
- **Text Decoder** est un modèle vision-to-text (comme GIT [6] ou blip2 [3]), modifié pour le format 3D, qui permet de générer les tokens et donc les IDs.
- **Beam Search**, ou la recherche en faisceau, renvoie les IDs des prédictions, ordonnés par score de probabilité.

Expérimentations

Les datasets choisis pour les expérimentations sont KITTi [1] et MulRAN [2], acquis par scans terrestres et organisés en séquences de milliers de sous-nuages de points géolocalisés. Chaque séquence contient des scènes revisitées donc des couples de nuages géométriquement proches. Notre proposition est évaluée face aux approches de l'état de l'art comme PointNetVLAD [4] et LoGG3D-NET [5], selon une comparaison par F1-score maximal de chaque séquence.

Pour une série de nuages, notre méthode propose plusieurs nuages candidats, ordonnés par probabilités décroissantes.
Les évaluations sont en cours et les premiers tests ont permis de vérifier la présence et le rang d'une revisite. Pour un nuage requête (en bleu, voir figure ci-contre), on affiche ici la prédiction la plus probable en rouge.
Les tests suivants consisteront à réaliser une classification binaire de chaque nuage pour déterminer un F1-score par séquence.



Exemple de prédiction correcte (en bleu, le nuage requête en rouge le nuage prédit la plus probable)

Perspectives

Les axes de recherche sont les suivants :

- Évaluer le comportement de la proposition, en changeant les datasets et les différents blocs de l'architecture.
- Ajouter l'association multi-modale terrestre-aérien
- Améliorer l'interface utilisateur, lier les adresses aux nuages

Informations

- Université Gustave Eiffel - ED MSTIC
- Laboratoire : LASTIG
- Direction: Valérie Gouet-Brunet
- Encadrant: Laurent Caraffa

Bibliographie

[1] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[2] Giseop Kim, Yeong Sang Park, Younghun Cho, Jinyong Jeong, and Ayoung Kim. Mulran: Multimodal range dataset for urban place recognition. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6246–6253. IEEE, 2020.

[3] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*, pages 19730–19742. PMLR, 2023.

[4] Mikaela Angelina Uy and Gim Hee Lee. PointNetVLAD: Deep point cloud based retrieval for large-scale place recognition. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4470–4479. IEEE.

[5] Kavisha Vidanapathirana, Milad Ramezani, Peyman Moghadam, Sridha Sridharan, and Clinton Fookes. LoGG3d-net: Locally guided global descriptor learning for 3d place recognition. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2215–2221.

[6] Jianfeng Wang, Zhengyuan Yang, Xiaowei Hu, Linjie Li, Kevin Lin, Zhe Gan, Zicheng Liu, Ce Liu, and Lijuan Wang. GIT: A generative image-to-text transformer for vision and language. *Transactions on Machine Learning Research*, 2022.

[7] Honghui Yang, Tong He, Jiaheng Liu, Hua Chen, Boxi Wu, Binbin Lin, Xiaofei He, and Wanli Ouyang. Gd-mae: generative decoder for mae pre-training on lidar point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9403–9414, 2023.

[8] Shengyao Zhuang, Houxing Ren, Linjun Shou, Jian Pei, Ming Gong, Guido Zuccon, and Daxin Jiang. Bridging the gap between indexing and retrieval for differentiable search index with query generation. *arXiv preprint arXiv:2206.10128*, 2022.