© IGN – LiDAR scan by terrestrial mobile mapping

© IGN – HD LiDAR scan by aerial mobile mapping

Thesis proposal

Point cloud based large-scale place recognition

Application to the prevention against fake news

Subject of the thesis

The thesis project focuses on 3D point cloud based large-scale place recognition, with the application of geolocation of 3D image data. Without any extra information of the initial position, geolocalazing image content relies on the indexing and retrieval of content similarities in a geolocalized reference. This thesis proposes to study this type of approach by exploiting 3D maps based on acquisition campaigns (in particular LiDAR) that are becoming mainstream thanks to high quality geometry reconstruction which makes them attractive, but also complex to handle given their volume and diversity.

Context

The fields of application of place recognition from images are numerous, we will deal here with the case of the geolocation of amateur video sequences as a certification tool for the prevention against fake news. Massively spread on social networks and on the web, amateur videos relaying information or an event are now very important, with among them content that is *fake news*, *i.e.* taken outside of its original context, to express bad or false information. To fight against this form of misinformation, several media, such as the French public television channel “France TV”, have set up a *fact checking unit*[1] of images and videos which analyzes, verifies and certifies these streams. This complex work is done by hand and would benefit from being automated by using artificial intelligence tools. The verification of geolocation was recognized as essential to best explain what is happening. It is in this collaborative context between IGN and France TV that we focus on this geolocation criterion with the desire to exploit the best georeferencing repositories of today to offer automatic large-scale geolocation solutions, which can, among other things, contribute to the *fact checking* of visual information.

Description of the thesis subject

Without a priori knowledge of an initial position, georeferencing image content, whether 2D, 2D+t, 3D or nD, is handled by indexing and retrieving content at large-scale: we start by looking for one or more similar places in terms of similarity of content in a set of similar nature (potentially large) which maps the environment and where the data is geolocated. The similar geolocated data found provides one or

[1]France TV fact checking unit (in French): https://www.francetvlab.fr/articles/les-revelateurs-de-francetv-la-cellule-de-fact-checking-video-et-images-de-france-televisions

more candidate locations. If necessary, it can be refined, for example by estimating the pose using a registration step.

With the increasing number of image-based learning datasets of places (e.g. Google Landmarks) and the enthusiasm for the automation of mobile systems (autonomous navigation, drones, etc.), the research on image-based localization, and its multiple variations (“landmark retrieval”, “place recognition”, etc.), has been growing rapidly [Pion et la, 2020], with different objectives in terms of location precision [Blettery et al, 2021] and of source of information exploited [Piasco et al, 2021]. Most of the approaches rely on 2D images, for which visual feature extraction has been a well-known process for decades. The exploitation of three-dimensional information describing the scene is less widespread, often starting from the assumption that with 3D information (LiDAR, SLAM, photogrammetry, etc.) an initial location is available (e.g. GPS), which reduces localization to a problem of registering 2D or 3D data in a 3D cloud [Uy et al, 2018] [Huang et al, 2021].

In this thesis, we have chosen to exploit this 3D information to respond to the problem of content geolocation, based on the principle that 1/ 3D data sources mapping the environment are more and more widespread, more and more precise with different natures (different LiDAR, SfM, photogrammetry, RGB-D sensors, etc.) and 2/ these geometric data provide robust representations to scene variations such as the change of viewpoint or the day time (night /day, season, etc.), which continue to penalize approaches based on the photometry of 2D images [Uy et al, 2018]. When it comes to searching in a large 3D reference volume without knowledge of an initial position, state-of-the-art techniques are generally grouped together under the name "Point Cloud Based Large-Scale Place Recognition", which rely on large-scale location recognition through similarity search in a large volume of reference 3D clouds, from a query 3D cloud.

Regardless of the data being manipulated, large-scale place recognition by similarity requires the description and structuring of query and baseline data. These two problems are addressed in the thesis, and we break them down into three research topics:

1. **Single-source description.** First, the problem of location search will be treated by considering a single 3D data source, one LiDAR cloud type, to study or improve the robustness of the best approaches according to the data and the volume. We will study how to describe these clouds with the goal of searching by similarity in the reference base. The state of the art on point cloud description is already vast, with handcrafted descriptors and learned solutions, spread in the fields of indexing 3D objects [Bold et al, 2019] and of registering 3D point clouds [Zhou et al, 2020][Huang et al, 2021]. When it comes to build descriptors for searching in a large 3D volume, the most recent techniques rely on deep learning of local and global descriptors extracted from the whole scene [Liu et al, 2019][Hui et al, 2021][Xia et al, 2021][Komorowski, 2021][Hui et al, 2022] and follow the 1st proposal on the subject, PointNetVLAD [Uy et al, 2018].

2. **Multi-source description.** If we consider as a query a sequence of images coming from a standard device (smartphone), the corresponding 3D point cloud can be obtained by Structure-from-Motion, and its nature will be different of the reference cloud. Moreover, the sources of 3D representations of the environment are now numerous and varied: different types of LiDAR (Riegl, Velodyne, aerial HD), SLAM, photogrammetry, etc. Exploiting the available modalities would bring robustness in the search for place while being exploitable on a larger scale. This is a widespread problem in remote sensing [Zhang et al, 2022], but still little addressed when it comes to 3D point clouds; the few approaches considering multi-source clouds rely heavily on the estimation of the transformation between clouds by registration [Huang et al, 2021], which would not be applicable here. Several ideas could be considered: treat each modality separately and merge the candidate responses a posteriori, or more ambitiously, learn a description that is invariant to the modality, what is called cross-domain description, as is the case with several approaches in remote sensing [Zhang et al, 2022] or in indexing heritage content as we have already studied in [Gominski et al, 2021].

3. **Large-scale indexing and retrieval.** Searching for a place by similarity of descriptors at large scale supposes to efficiently browse a large volume of multidimensional descriptors. Surprisingly, the current best approaches to 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] are evaluated on a subset of the Oxford Robotcar dataset [Maddern et al, 2016] which corresponds to a LiDAR coverage of some

10 tens of km and 100M pre-processed points; comparatively, land mobile mapping acquisitions made in Paris by the IGN cover a length of 200 km for 70 billion raw points. If these works have focused on the richness and discrimination of the proposed descriptors and not on their structuring for the fastest possible research, it seems essential to consider scaling up, in terms of research time but also of robustness of the descriptions in relation to the volume. The most efficient recent indexing techniques are based on approximate retrieval using hash tables, already well established in image indexing [Rodrigues et al, 2020] and which have become popular in remote sensing [Reato et al, 2019]. Depending on the characteristics of the descriptors learnt, it will therefore be necessary to set up a representation approach with binary hash codes, easily distributable and allowing an efficient search in the descriptors of point clouds. Among the various approaches considered, we will consider hierarchical approaches [Zhang et al, 2021-a], which currently seem to provide the best results.

Datasets considered

It is necessary to consider large-scale datasets with ground truth, first to qualitatively and quantitatively validate the proposed approaches, then to provide ground truth to approaches based on learning. The work will be based on international public single-source and multi-source benchmarks in the field (e.g. [Maddern et al, 2016] [Zhang et al, 2021-b]) but also on IGN data sets such as the 2D and 3D mapping of all of Paris by terrestrial mobile mapping[2] and the aerial HD LiDAR campaign being acquired nationwide[3].

Candidate profile

Bac+5 in computer science, applied mathematics or geomatics (master or engineering school). A good background in machine learning is required, and a knowledge on 3D computer vision or image indexing will be appreciated. The successful candidate must have good programming skills (Python, C/C++).

Although fluency in French is not required, fluency in English is necessary. Curiosity, open-mindedness, creativity, perseverance and the ability to work in a team are also key personal skills in demand.

Only students from the European Union, the United Kingdom or Switzerland are eligible for this thesis project.

Organization

Start: last quarter of 2022

Place: the thesis will be carried out in Paris area at the LaSTIG laboratory, located in Saint-Mandé (73 avenue de Paris, Saint-Mandé metro, line 1) in the premises of the IGN. The doctoral student will be attached to the MSTIC Doctoral School (ED 532).

The French mapping agency IGN (National Institute for Geographic and Forest Information) is a public administrative establishment attached to the French Ministry of Ecological Transition; it is the national reference operator for mapping the French territory. The LaSTIG[4] Laboratory in Sciences and Technologies of Geographic Information for the smart city and sustainable territories, is a joint research unit attached to the Gustave Eiffel University, the IGN and the School of Engineering of the city of Paris (EIVP). It is a unique research structure in France and even in Europe, bringing together around 80 researchers, who cover the entire life cycle of geographic or spatial data, from its acquisition to its visualization, including its modeling, integration and analysis; among them about thirty researchers work in image analysis, computer vision, machine learning, photogrammetry and remote sensing.

LaSTIG researchers can be involved in the teaching activities of the IGN engineering school, the ENSG (Ecole Nationale des Sciences Géographiques), which offers access to undergraduate and

[2]Stereopolis IGN land mobile mapping: https://www.ign.fr/institut/donnees-collectees-par-le-vehicule-stereopolis
[3]IGN HD LiDAR: https://www.ign.fr/institut/lidar-hd-vers-une-nouvelle-cartographie-3d-du-territoire
[4]LaSTIG website: https://www.umr-lastig.fr/

graduate students with excellent quality in fields related to geographic information sciences: geodesy, photogrammetry, computer vision, remote sensing, spatial analysis, cartography, etc.

How to apply

Before March 28, 2022, please send both contacts in a single PDF file the following documents:

- A detailed CV
- A topic-focused cover letter
- Grades and ranks over the last 3 years of study
- The contact details of 2 referents who can recommend you

Contacts

- Laurent Caraffa – Laurent.Caraffa@ign.fr
  Researcher at LaSTIG (thesis supervisor), IGN, Gustave Eiffel University
- Valérie Gouet-Brunet – Valerie.Gouet@ign.fr
  Research director at LaSTIG (director of the thesis), IGN, Gustave Eiffel University

References

[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, p. 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, p.p. 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, p. 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 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 ConditionsInternational Journal of Computer Vision, Springer Verlag, 129 (1), pp.185- 202, 2021.

[Pion et la, 2020] Noé Pion, Martin Humenberger, Gabriela Csurka, Yohann 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, p. 276-280, Feb. 2019.

[Rodrigues et al, 2020] Rodrigues, J., Cristo, M. & Colonna, JG Deep hashing for multi-label image retrieval: a survey. Artif Intel 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), p. 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.