PhD Thesis proposal

Indexing and retrieval of visual contents in 3D point clouds at large scale

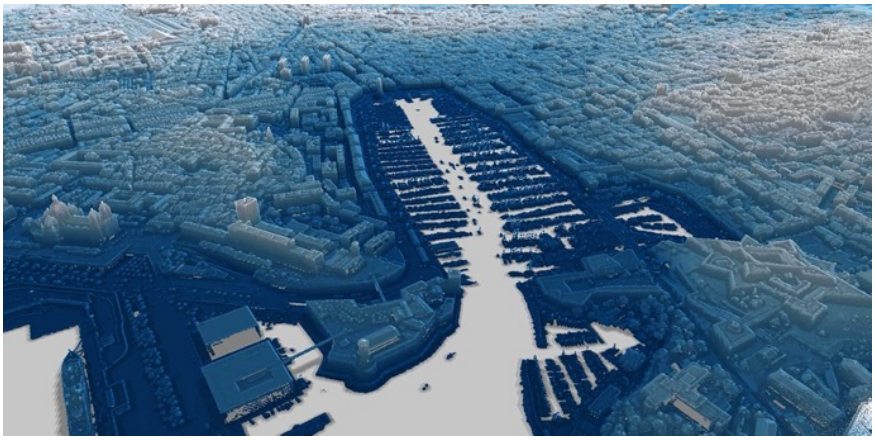
Application to spatialization

At a glance

The thesis project focuses on the spatialization of visual contents (both image and video contents) by the exploitation of 3D references at large scale. Without any a priori about geolocation, the problem is tackled by the retrieval of the most similar elements in the geolocalized reference. As visual content, we consider old photographs and footages made available from cultural institutions, and as 3D reference we exploit LiDAR data mapping the French territory, made available at the country scale by the French mapping agency (IGN). This PhD thesis has the ambition to address two challenging scientific problems: on the one hand, the description, matching and indexing of 2D(+t) and 3D data in a multi-date context where the scene has evolved over time, and on the other hand, the fast retrieval in very large volumes of data. The work will be carried out within the framework of the multidisciplinary project AGAPE, which addresses the discoverability and investigation in spatial iconographic heritage, and gathers seven leading partners specialized in visual and multimodal AI, Multimedia and Human-Computer Interaction as well as in Archives, History and Media.

© Archives nationales, Fonds Lapie

© IGN, Photothèque Nationale, Durandeau 1947



© IGN – HD LiDAR scan by aerial mobile mapping (2024)

Fig. 1: Marseille harbor at different periods (from 1947 to nowadays).
On the left, old aerial oblique imagery; on the right, a recent high-resolution LiDAR scan

Keywords

Computer Vision, Artificial Intelligence, Indexing and Retrieval, Vision Languages Models, Image analysis, 3D Point Clouds, Big Data, Geolocalization, Cultural Heritage.

Context

AGAPE (2025-2028) is a French project which concerns the discoverability and investigation in spatial heritage iconographic contents, with a focus both on their multimodal analysis, linking and confrontation, and on their joint positioning in 3D context to facilitate visualization and user interaction. Within this project, the PhD student will contribute on the linking and spatialization of collections of image and video contents, for their integration in the 3D environment. The consortium brings together seven leading partners, among them four specialized in Computer Science, Artificial Intelligence, Multimedia and Human-Computer Interaction (Labs. LASTIG, LIRIS, IRISA/Linkmedia and Inria/ILDA) and three other ones in charge of the exploitation of the tools proposed in three different domains: archivists (Archives nationales de France), historians (Lab. ACP) and journalists (France Televisions).

Subject

The PhD student will study solutions to organize visual collections (old photographs and footages) by analysis of their content. By organization, we mean their linking according to their spatial proximity, in order to gather those that document the same area, and their spatialization in order to restitute them in their 3D environment. Classically, this problem is tackled by image retrieval in a dataset of geolocalized images, which links similar contents visually, and propagates location through them [Pion et al, 2020]. More originally here, we will address the problem by selecting a global spatialized 3D reference, and positioning the 2D contents within it, through retrieval in this reference. We choose to consider 3D point clouds as reference, mainly LiDAR, based on the observations that 1/ 3D data sources mapping the environment are more and more widespread, more and more precise (IGN LiDAR HD mapping) and 2/ these geometric data provide geolocalized representations robust 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. Retrieving 2D(+t) contents in a 3D dataset does, however, involve some unresolved scientific bottlenecks, which will be at the heart of the work to be carried out:

- **Cross-domain indexing:** it requires finding representations that allow to make links between 2D(+t) and 3D. Literature is rich concerning indexing and retrieval in one domain, such as images [Pion et al, 2020] [Blettery et al, 2023] or 3D point clouds [Uy et al, 2018] [Komorowski, 2021] [Hui et al, 2022] [Vidanapathirana et al, 2022] [Zhang et al, 2024]. Multiple solutions for 2D-3D matching also exist [Li et al, 2021] [Qin et al, 2022], under the topic of registration, but are not applicable to retrieval at large scale because they suppose the knowledge of the 2D content geolocation to register it with the 3D. As with Visual Language Models which align image and text contents, we will rather consider methods that provide a common embedding between 2D and 3D clouds [Xie et al, 2023] [Xue et al, 2023] [Sohail et al, 2025], exploitable as index for retrieval, but by focusing on the difficulty of aligning such contents in a multi-date context where the scene has evolved over time.
- **Retrieval at large scale** (at least city-wide): our work will take as starting point the work already carried out in LASTIG on 3D retrieval in point cloud datasets [Zede et al, 2025] inspired from Differentiable Search Index [Zhuang et al, 2022], GD-MAE Sparse Pyramid Transformers [Yang et al, 2023] and Beam search, which drastically accelerate retrieval in large datasets. To improve the scalability of these approaches, the use of the Mamba architecture [Gu et al, 2024], both for point cloud encoding via MambaPoint [Cheng et al, 2024] and at the core of the Differentiable Search Index, can be explored, as well as Mixture of Experts architectures that offer promising directions, easier to train and to update [Cai et al, 2025].

As a member of the AGAPE consortium, the PhD student is also in charge of interacting with its members (researchers, post-docs, PhD students, interns, archivists, historians and journalists), and in particular of participating in the integration of the geolocalized contents in the 3D environment with partners, and in the evaluation of his/her proposals for the applicative scenarios of the project with the archivists, historians and journalists. He will also contribute to the integration of the developed algorithms into the consortium computing infrastructures to ensure their deployment and scalability. AGAPE is also actively involved in open science, all results obtained and solutions developed will be published in high-level conferences and journals, and deposited in open access.

Datasets considered

Through the AGAPE's consortium, the participation of IGN (French mapping agency) and open data available in the domain of cultural heritage, the PhD student will benefit from various and large datasets: old collections of image and video contents representing the French territory, at the aerial and terrestrial

levels (over 70,000 images); IGN 3D LiDAR point clouds at large scale, of whole Paris by terrestrial mobile mapping and by the aerial HD LiDAR campaign being acquired nationwide.

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 and experiences in 3D computer vision or image indexing will be highly appreciated. The successful candidate must have good programming skills (Python, C/C++). Knowledge of software engineering tools and practices such as Docker, Kubernetes, etc., is a plus.

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

Organization

Start: flexible, ideally last quarter 2025.

Funding: fully funded (3-year doctoral contract and missions abroad).

Place: the thesis will be carried out in Great Paris area at the LASTIG laboratory, located on the campus of the Gustave Eiffel University in Champs-sur-Marne. The doctoral student will be attached to the MSTIC Doctoral School (ED 532).

The LASTIG 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.

The members of the LASTIG work in close collaboration with the French mapping agency (IGN, National Institute for Geographic and Forest Information), which, as public administrative establishment attached to the French Ministry of Ecological Transition, is the national reference operator for mapping the French territory. LASTIG researchers and PhD students 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 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 July 14, 2025, please send to 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

Candidatures which do not respect these instructions will not be considered.

Auditions will be conducted during period July 15-23; decision released no later than July 25.

Contacts

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

References

[Blettery et al, 2023] Emile Blettery and Valérie Gouet-Brunet, *Re-ranking Image Retrieval in Challenging Geographical Iconographic Heritage Collections*, 20th International Conference on Content-Based Multimedia Indexing, Sep 2023, Orléans, France.

[Cai et al, 2025] W. Cai, J. Jiang, F. Wang, J. Tang, S. Kim and J. Huang, *A Survey on Mixture of Experts in Large Language Models*, in IEEE Transactions on Knowledge and Data Engineering, pp. 1-20, March 2025.

[Cheng et al, 2024] R. Cheng, Z. Xu, Y. Zhang, B. Yang, *PointMamba: A Simple State Space Model for Point Cloud Analysis*, in arXiv preprint arXiv:2402.05352, 2024.

[Gu et al, 2024] A. Gu, K. Dao, T. Lin, S. Rekhi, J. Yao, A. Rush, T. Ma, *Transformers are SSMs: Generalized Models and Efficient Algorithms Through Structured State Space Duality*, in International Conference on Learning Representations (ICLR), 2024.

[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.

[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.

[Li et al, 2021] Jiaxin Li, Gim Hee Lee, *DeepI2P: Image-to-Point Cloud Registration via Deep Classification*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 15960-15969.

[Pion et al, 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.

[Qin et al, 2022] Z. Qin, H. Yu, C. Wang, Y. Guo, Y. Peng, K. Xu, *Geometric Transformer for Fast and Robust Point Cloud Registration*, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2022, pp. 11143–11152.

[Sohail et al, 2025] Shahab Saquib Sohail, Yassine Himeur, Hamza Kheddar, Abbes Amira, Fodil Fadli, Shadi Atalla, Abigail Copiaco, Wathiq Mansoor, *Advancing 3D point cloud understanding through deep transfer learning: A comprehensive survey*, Information Fusion, Volume 113, January 2025.

[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.

[Vidanapathirana et al, 2022] 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. IEEE, 2022.

[Xie et al, 2023] Yichen Xie, Chenfeng Xu, Marie-Julie Rakotosaona, Patrick Rim, Federico Tombari, Kurt Keutzer, Masayoshi Tomizuka, Wei Zhan, *SparseFusion: Fusing Multi-Modal Sparse Representations for Multi-Sensor 3D Object Detection*, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 17591-17602

[Xue et al, 2023] Le Xue, Mingfei Gao, Chen Xing, Roberto Martín-Martín, Jiajun Wu, Caiming Xiong, Ran Xu, Juan Carlos Niebles, Silvio Savarese, *ULIP: Learning a Unified Representation of Language, Images, and Point Clouds for 3D Understanding*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 1179-1189.

[Zhang et al, 2024] Yongjun Zhang, Pengcheng Shi, and Jiayuan Li, *LiDAR-based place recognition for autonomous driving: A survey*, ACM Computing Surveys, 57(4):1–36, 2024.

[Zede et al, 2025] Chahine-Nicolas Zede, Laurent Caraffa and Valerie Gouet-Brunet, *Fast 3D point clouds retrieval for Large-scale 3D Place Recognition*, ORASIS, journées francophones des jeunes chercheurs en vision par ordinateur, Jun 2025, Le Croisic, France.