PhD Thesis proposal

Scalable indexing and retrieval
in multimedia and geospatial contents

At a glance

Within the DALEAS project, the PhD will tackle the core challenge of content-based indexing and retrieval in multimedia contents at large scale. The multimedia contents considered, such as images, text, and 3D point clouds, illustrate or document the territory (a city, street, monument, etc.), as illustrated in Fig. 1. The research will consist in designing efficient, flexible representations and indexing strategies that enable fast and accurate retrieval, depending on the data available in the query as well as in the georeferenced database queried. Here, retrieval will enable the identification of documents linked to a common geographical area, with the objective of geolocating the query. Building on previous work on LiDAR-to-text representations performed in LaSTIG, and on Multimodal Large Language Models in general, the candidate will propose fusion and alignment strategies to integrate heterogeneous data. Embedded in the Apache Spark™ infrastructure, this work will advance large-scale search in multimedia collections, with application to geospatial analysis of social network media in partnership with France Télévisions, for the detection of misinformation.

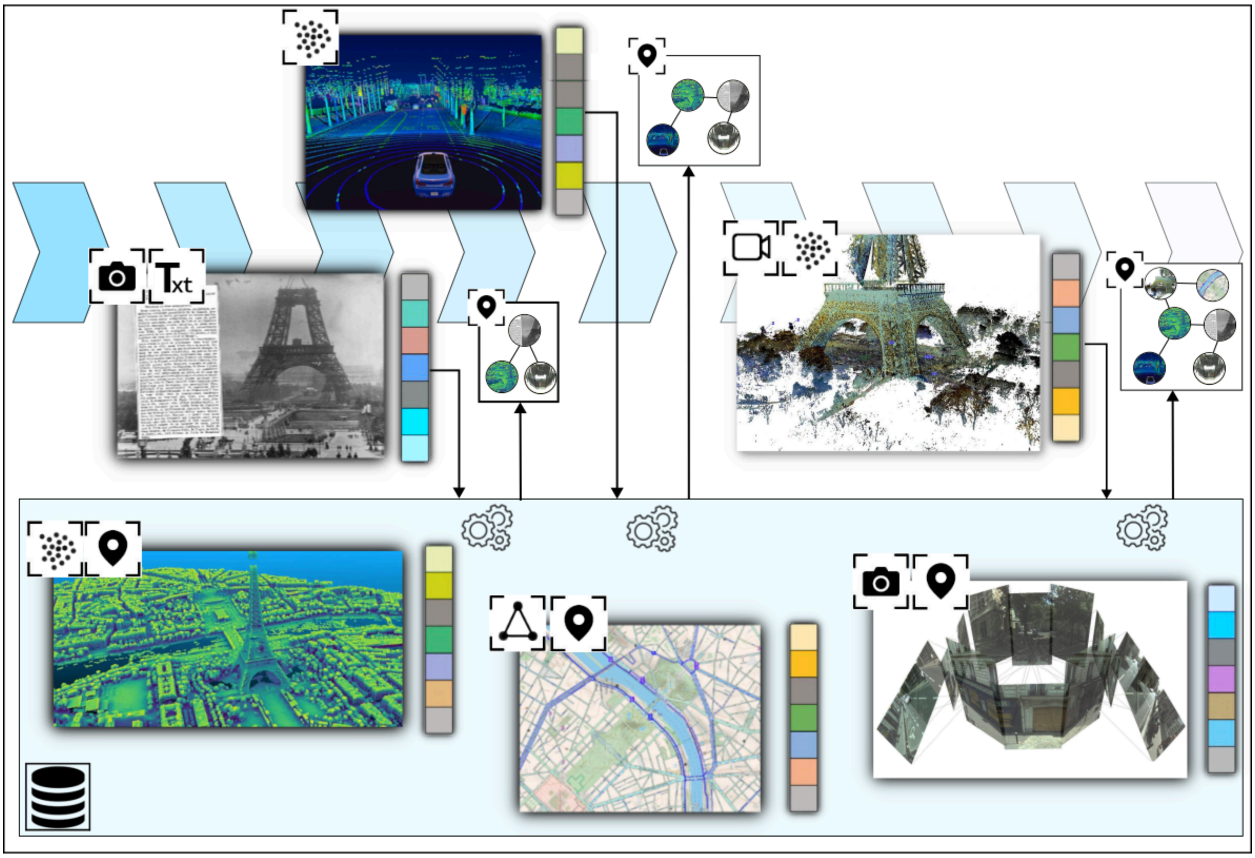

Fig 1 : Multimodal data on the area of Paris (archival documents, maps, 3D point clouds from aerial and mobile mapping) that are streamed, aligned and connected to geolocalized sources into a unified processing infrastructure.

Keywords

Information retrieval, Indexing and Retrieval, Machine Learning, Multimodal Large Language Models, 3D Point Clouds, Geolocalization, Big Data, Apache Spark.

Context

DALEAS (2026-2028) is a French research project that aims to design an infrastructure capable of analyzing very large volumes of multimedia data (images, videos, text, 3D, etc.) in real time, in order to automatically detect data that are misaligned with a reference database. By relying on trusted reference data, such as those of the French mapping agency (IGN), and on the FAIR4RS principles [Barker et al., 2022], DALEAS aspires to provide a collaborative, flexible, transparent, and sustainable tool, reusable in any context requiring large-scale geospatial data stream analysis. This proof of concept, implemented with the Apache Spark™ framework, is designed for seamless scalability. It will provide a foundation for new research on real-time anomaly detection, centered on large-scale multimodal indexing and retrieval — the ability to rapidly search massive databases using potentially complex criteria and various types of data. The consortium brings together IJCLab and LaSTIG, two partners with advanced and complementary expertise in large-scale, particularly spatial, data processing. They will deliver innovative software and algorithmic solutions for massive multimedia data processing, building on the Fink architecture [Möller et al, 2021], under the umbrella of transparency, adaptability, reproducibility, and accessibility. Through the design of this infrastructure, DALEAS seeks to address the needs of diverse sectors: we have the ambition to showcase its capabilities for combating disinformation with France Télévisions, investigation support with la Gendarmerie nationale, digital twins updating with IGN, and space observation with IJCLab.

The PhD will contribute to the design and evaluation of scalable multimedia indexing and retrieval strategies within the DALEAS framework. In particular, he will support the use case on disinformation detection through the geolocalization of data streams by content analysis and alignment, based on the retrieval of documents characterizing the same location as the query one, leveraging the available modalities — text, images, and 3D point clouds (LiDAR) in our case.

Subject

Indexing and retrieval in a multimedia collection rely both on an indexing step, i.e. on the **description** of the query content and of each item of the collection, and on a **retrieval** step, i.e. on the structuring of the descriptions for fast comparison in the collection. When considering several modalities in the query or in the collection queried, it is necessary to introduce a step of **fusion** of the information, which can be achieved in multiple ways, at the level of the description (early fusion, e.g. [Gadzicki et al., 2020]), of the retrieval step itself (intermediate fusion, e.g. [Valenzuela et al., 2014]), of the responses returned (late fusion, e.g. [Neshov, 2013; Ye et al., 2012]), or even sequentially when modalities are used in a hierarchical manner to filter and refine candidate sets (sequential fusion, e.g. [Mai et al, 2019]). If the query's modality is not the same as the one of the items in the collection, it is necessary to **align** the modalities; furthermore, alignment is also possible to solve the fusion problem.

The thesis focuses on the fusion and alignment problems dedicated to retrieval. We draw inspiration from recent advances in Multimodal Large Language Models (MLLMs [Caffagni et al., 2024; Zhang et al., 2024]), such as TEAL [Yang et al., 2024], where the interactions between any modalities are treated as a token sequence and learnt in a joint embedding space. Before this, the team explored retrieval in LiDAR datasets by leveraging generative Image-to-Text base model [Wang et al., 2022] to generate language representations of 3D point clouds for large-scale place recognition [Zede et al., 2025]. Building on these results and on the last generation of MLLMs, the objective of this research is to generalize the concept to a broader and variable range of modalities, including images, text, and/or 3D data, both in the query and in the database queried, by developing a unified tokenization strategy and scalable indexing methods capable of supporting large-scale retrieval in multimedia collections.

The concepts developed will be integrated into the Apache Spark™–based DALEAS infrastructure, and will be applied to **geolocalization**, by being able to retrieve georeferenced documents similar to the one of the query. This step of geolocalization will be exploited to geolocalize data streams, with the final objective of **stream fact-checking**, with partner France Télévisions. Depending on the multimedia data available, it is highly probable that the candidate will have to study, and compare different strategies of fusion - joint fusion (i.e. early, intermediate, late) or sequential fusion - optimized for this downstream task and sufficiently flexible to accept various modalities, rather than committing to a single approach upfront.

Datasets considered

Within the DALEAS consortium, the PhD student will have access to a variety of large-scale and heterogeneous datasets combining multimedia and geospatial information. These include open geospatial data and reference collections provided by project partners, such as the LiDAR HD dataset from the IGN, a nationwide high-density aerial point cloud acquisition (10 points/m²), or stereo image couples from the IGN Stereopolis mobile mapping system [Paparoditis et al., 2012], as well as publicly available resources defined by the project use case, such as social media like X (formerly Twitter).

Candidate profile

Bac+5 in computer science, applied mathematics or geomatics (master or engineering school). A good background in machine learning is required, and knowledge and experiences in information retrieval, image indexing, retrieval or computer vision 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, Apache Spark™ etc., is a significant plus.

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

Organization

Start: flexible, ideally first quarter 2026.

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

Place: the thesis will be carried out in the Great Paris area at 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 (National Institute for Geographic and Forest Information - French mapping agency), and the EIVP (School of Engineering of the city of Paris). 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 30 researchers work in image analysis, computer vision, machine learning, and remote sensing.

The members of the LASTIG work in close collaboration with the IGN, which, as a 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 Geodata Paris (ex-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 November 21th, 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.

Applications will be accepted until November 21st, with auditions conducted from November 24th to December 5th, and the decision released by December 10th.

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

[Barker et al., 2022] M. Barker, N.P. Chue Hong, D.S. Katz, et al. Introducing the FAIR Principles for research software. Scientific Data, 9:622, 2022. https://doi.org/10.1038/s41597-022-01710-x

[Caffagni et al., 2024] Davide Caffagni, Federico Cocchi, Luca Barsellotti, Nicholas Moratelli, Sara Sarto, Lorenzo Baraldi, Marcella Cornia, et Rita Cucchiara. The (R)Evolution of Multimodal Large Language Models: A Survey. In Findings of the Association for Computational Linguistics: ACL 2024, pages 13590–13618, Bangkok, Thailand, 2024.

[Gadzicki et al, 2020] K. Gadzicki, R. Khamsehashari and C. Zetzsche, Early vs Late Fusion in Multimodal Convolutional Neural Networks, 2020 IEEE 23rd International Conference on Information Fusion (FUSION), Rustenburg, South Africa, 2020, pp. 1-6, doi: 10.23919/FUSION45008.2020.9190246.

[Mai et al, 2019] Sijie Mai, Haifeng Hu, and Songlong Xing. Divide, Conquer and Combine: Hierarchical Feature Fusion Network with Local and Global Perspectives for Multimodal Affective Computing. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 481–492, Florence, Italy, 2019.

[Möller et al, 2021] Anais Möller, Julien Peloton, Emille E. O. Ishida, et al. Fink, a new generation of broker for the LSST community. Monthly Notices of the Royal Astronomical Society, 501(3):3272–3288, 2021.

[Neshov 2013]. Comparison on late fusion methods of low level features for content based image retrieval. In : International Conference on Artificial Neural Networks. Berlin, Heidelberg : Springer Berlin Heidelberg, 2013. p. 619-627.

[Paparoditis et al., 2012] N. Paparoditis, J.-P. Papelard, B. Cannelle, A. Devaux, B. Soheilian, N. David, et E. Houzay. Stereopolis II: A multi-purpose and multi-sensor 3D mobile mapping system for street visualisation and 3D metrology. Revue française de photogrammétrie et de télédétection, vol. 200, no 1, pp. 69–79, 2012.

[Valenzuela et al, 2014] Ricardo E. Gonzalez Valenzuela, Neelanjan Bhowmik, Valerie Gouet-Brunet, and Helio Pedrini. Efficient fusion of multidimensional descriptors for image retrieval. In International Conference on Image Processing, pages n.n., 2014.

[Wang et al., 2022] 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 (accepted for publication), arXiv preprint arXiv:2205.14100, 2022.

[Yang et al., 2024] Zhen Yang, Yingxue Zhang, Fandong Meng, and Jie Zhou. TEAL: Tokenize and Embed ALL for Multi-modal Large Language Models. arXiv preprint arXiv:2311.04589, 2024.

[Ye et al, 2012] Guangnan Ye, Dong Liu, I-Hong Jhuo, and Shih-Fu Chang. Robust late fusion with rank minimization. In Computer Vision and Pattern Recognition, pages 3021–3028, 2012.

[Zede et al., 2025] Chahine-Nicolas Zede, Laurent Caraffa, and Valérie Gouet-Brunet. DSI-3D : Differentiable Search Index For Point Clouds Retrieval. In 22th IEEE International Conference on Content-Based Multimedia Indexing (CBMI'25), p. 1-7, Dublin, Ireland, October 2025.

[Zhang et al., 2024] Duzhen Zhang, Yahan Yu, Jiahua Dong, Chenxing Li, Dan Su, Chenhui Chu, et Dong Yu. MM-LLMs: Recent Advances in MultiModal Large Language Models. In Findings of the Association for Computational Linguistics: ACL 2024, pages 12401–12430, Bangkok, Thailand, 2024.