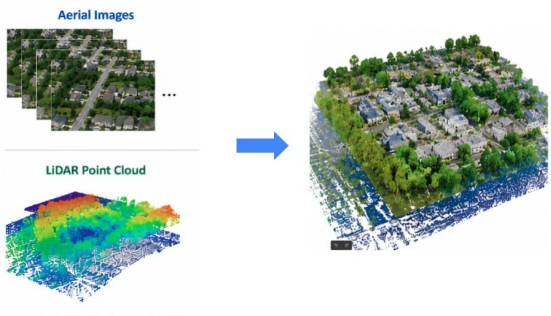
PhD position 2026

LiDAR–Image Fusion for Accurate and Domain-Adaptive Multi-View Stereo Reconstruction using Transformer-based Architectures

UGE/LASTIG/ENSG/IGN



Fusion images and LiDAR

3D point cloud

1 Keywords

Computer Vision, Photogrammetry, denes matching, multi view stereo, deep learning, multi-model

2 Contexte

Traditional 3D reconstruction methods based on stereo dense matching or Multi-View Stereo (MVS) reconstruction rely solely on photogrammetry and often fail in areas with low texture, specular surfaces, or complex geometries [6,8]. Meanwhile, LiDAR systems produce dense and accurate point clouds but lack continuous radiometric information, and the acquisition is expensive. Traditional methods struggle in difficult situations, for example, with low texture and thin objects [13]. With the development of deep learning, deep learning stereo dense matching or deep MVS methods have revolutionized 3D reconstruction by learning implicit stereo correspondence estimation from large datasets. But deep learning stereo dense matching still needs a fusion disparity map to a point cloud, and this degrades the advantage of the learning method, so deep MVS provides an end-to-end method for reconstruction. However, these methods remain limited in two key aspects: (1) Geometric accuracy — especially in textureless or repetitive regions; (2) Domain generalization — models trained on specific datasets fail to perform on new domains. LiDAR, with its dense and metrically precise 3D measurements, offers an ideal complementary source of geometric information [12].

3 Introduction and goals

3.1 State of the art

3D reconstruction is an important topic both for photogrammetry and computer vision. With the development of deep learning, learning methods outperform traditional methods. Stereo dense matching only uses two-view images; in real 3D reconstruction applications, large overlap can be achieved, and multi-view stereo can improve the robustness and ambiguity. Recent research on MVS (MVS-Net [14], CasMVSNet [11], TransMVSNet [2], PatchmatchNet [10], GeoMVSNet [16]) has advanced 3D reconstruction through deep learning.

Deep learning based MVS is a widely explored topic in computer vision, because there are many datasets for example, DTU [3], Tanks and Temples [4], ETH3D [7], and BlendedMVS [15]. So there are many types of learning MVS methods, which can be categorized: (1) From cost-volume MVS to cascade; (2) PatchMatch and iterative methods; (3) Transformers and global context for improved matching method; (4) Geometry-aware based method; (5) Self-supervision based method; (6) Cross-modal fusion; (7) Large-scale application.

Concurrently, Transformer-based models have shown superior feature representation and matching capabilities [9]. Transformers bring significant advantages to deep learning MVS by modeling global context and long-range dependencies that CNNs cannot capture [1, 2]. Through self- and cross-view attention, they aggregate information from multiple images, improving correspondence estimation in textureless, reflective, or occluded areas. Their ability to jointly reason about geometry and appearance enhances metric accuracy, while their global feature modeling provides better domain generalization across different environments. Moreover, Transformers offer a flexible architecture that can fuse heterogeneous data such as LiDAR and imagery, making them ideal for accurate and domain-adaptive 3D reconstruction.

3.2 Goals

However, few studies have integrated LiDAR data to guide MVS learning or explored cross-domain adaptation [5]. The objective of the project is (1) to explore the fusion of LiDAR and image in the MVS framework using the Transformer. (2) domain adaptation, because different types of data influence the performance a lot, but it is impossible to access the training data for every application, so exploring the domain adaptation is also important for real applications.

1. **Benchmark Dataset Preparation**: IGN is the national geospatial data provider in France, offering nationwide aerial and satellite imagery, as well as the LiDAR HD dataset with a density of nearly 10 points/m^2 covering most of the country. Assuming that the imagery and LiDAR point clouds are accurately co-registered, the first objective of the project will be to generate high-quality benchmark datasets from these multimodal geospatial data sources.

2. **Transformer-based MVS with LiDAR Fusion**: The main objective of the PhD project is to develop a transformer-based framework for LiDAR-image fusion in multi-view stereo (MVS) reconstruction. Existing state-of-the-art methods will first be evaluated on the benchmark dataset. The project will then investigate how to integrate 3D LiDAR features into the MVS network in order to improve the robustness and accuracy of 3D reconstruction.

3. **Domain Adaptation**: Domain adaptation remains a major challenge in photogrammetry applications. Deep learning models trained on data from one city often suffer significant performance degradation when applied to another city, and the gap becomes even larger when transferring across different sensor types. Considering the availability of both aerial and satellite imagery, this project will explore domain adaptation strategies between these two sensing modalities to improve the generalization capability of the proposed methods.

4 Organization

Duration: The doctoral contract, is for a **3 years** period, and may or may not include teaching tasks, depending on the candidate's profile and preference.

Workplace:LASTIG Lab, Geodata Paris, Gustave Eiffel University, Champs-sur-Marne (RER A, station Noisy-Champs).

IGN (French Mapping Agency) is a Public Administrative Institution part of the French Ministry for Ecology and Sustainable Development. IGN is the national reference operator for the mapping of the territory; in particular, the agency is currently in charge of the 3D mapping program of France with LiDAR HD. The LASTIG is one of the research laboratories of IGN, attached to the Geodata Paris (ex-ENSG, Ecole Nationale des Sciences Géographiques), and Gustave Eiffel University (UGE) in Grand Paris area.

5 Candidate profile

Only students who are citizens of the European Union, the United Kingdom, or Switzerland are eligible. The candidate should hold a Master's degree in computer science, robotic or computer vision (master or engineering school); good knowledge in image, 3D data processing, and deep learning, as well as strong skills in programming (e.g. Python), knowing C/C++ is highly recommended. Good interpersonal skills, motivation for research and teamwork, initiative, writing skills, and proficiency in English are required.

6 Application

Send an email to the contacts below **in a single PDF file** before **June 7, 2026**:

- CV
- motivation letter
- 2 recommendation letters, or persons to contact
- Transcript of grades from the last two years of study

7 Contact

Bruno VALLET, senior researcher, LASTIG: bruno.vallet@ign.fr
Ewelina RUPNIK, researcher, LASTIG : ewelina.rupnik@ign.fr
Teng WU, researcher, LASTIG : teng.wu@ign.fr

References

[1] Chenjie Cao, Xinlin Ren, and Yanwei Fu. Mvsformer++: Revealing the devil in transformer's details for multi-view stereo. *arXiv preprint arXiv:2401.11673*, 2024.

[2] Yikang Ding, Wentao Yuan, Qingtian Zhu, Haotian Zhang, Xiangyue Liu, Yuanjiang Wang, and Xiao Liu. Transmvsnet: Global context-aware multi-view stereo network with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8585–8594, 2022.

[3] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engil Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 406–413. IEEE, 2014.

[4] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)*, 36(4):1–13, 2017.

[5] Matteo Poggi, Andrea Conti, and Stefano Mattoccia. Multi-view guided multi-view stereo. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8391–8398. IEEE, 2022.

[6] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1):7–42, 2002.

[7] Thomas Schöps, Johannes L. Schönberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, and Andreas Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[8] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, volume 1, pages 519–528. IEEE, 2006.

[9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[10] Fangjinhua Wang, Silvano Galliani, Christoph Vogel, Pablo Speciale, and Marc Pollefeys. Patchmatchnet: Learned multi-view patchmatch stereo. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14194–14203, 2021.

[11] Junkai Wang, Dazhong Ma, Qingchen Wang, and Jie Wang. Csa-mvsnet: a cross-scale attention based multi-view stereo method with cascade structure. *IEEE Transactions on Consumer Electronics*, 2025.

[12] Teng Wu, Bruno Vallet, and Marc Pierrot-Deseilligny. Psmnet-fusionx3: Lidar-guided deep learning stereo dense matching on aerial images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6527–6536, 2023.

[13] Teng Wu, Bruno Vallet, Marc Pierrot-Deseilligny, and Ewelina Rupnik. An evaluation of deep learning based stereo dense matching dataset shift from aerial images and a large scale stereo dataset. *International Journal of Applied Earth Observation and Geoinformation*, 128:103715, 2024.

[14] Yao Yao, Zixin Luo, Shiwei Li, Tian Fang, and Long Quan. Mvsnet: Depth inference for unstructured multi-view stereo. In *Proceedings of the European conference on computer vision (ECCV)*, pages 767–783, 2018.

[15] Yao Yao, Zixin Luo, Shiwei Li, Jingyang Zhang, Yufan Ren, Lei Zhou, Tian Fang, and Long Quan. Blendedmvs: A large-scale dataset for generalized multi-view stereo networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1790–1799, 2020.

[16] Zhe Zhang, Rui Peng, Yuxi Hu, and Ronggang Wang. Geomvsnet: Learning multi-view stereo with geometry perception. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 21508–21518, 2023.