Master 2 Research Internship Proposal

Evaluation of State-of-the-Art Feed-Forward 3D Reconstruction Methods for Mobile Robotics Applications

Context

Recent advances in computer vision have introduced powerful feed-forward transformer-based approaches that are revolutionizing 3D scene reconstruction and camera pose estimation. Two notable recent publications, VGGT (Visual Geometry Grounded Transformer) and MapAnything, present breakthrough methods that can predict camera parameters, depth maps, and 3D point clouds from multiple images in under one second. These approaches represent a significant paradigm shift from traditional optimization-based methods like Bundle Adjustment and Structure-from-Motion towards end-to-end neural networks. Both methods have open-source implementations available on GitHub, making them accessible for research and practical applications.

Research Objectives

This internship aims to evaluate and benchmark these state-of-the-art methods on real-world mobile robotics scenarios. The research will explore several important applications in the field of autonomous navigation and robotic perception.

One key aspect of the work will involve assessing how well these methods can relocalize a mobile robot within a pre-existing set of images. This is a fundamental capability for autonomous systems that need to understand their position in a known environment. The evaluation will consider various conditions including different lighting situations, partial occlusions, and other challenging scenarios commonly encountered in real-world deployments.

Another important direction will be the estimation of camera orientation parameters, particularly roll, pitch, and yaw angles for cameras mounted on mobile vehicles. Understanding the precise orientation of onboard cameras is crucial for many navigation and perception tasks. The research will investigate both the accuracy of these estimates and their potential for real-time applications in autonomous navigation systems.

The internship will also explore the possibility of using these feed-forward methods for extrinsic calibration of multi-camera systems mounted on mobile robots. Traditional calibration approaches often require specific calibration patterns or manual interventions. This research will investigate whether modern neural approaches can provide automatic, robust calibration solutions for various camera configurations including stereo rigs and surround-view systems.

Expected Outcomes

The internship should result in a comprehensive evaluation of VGGT and MapAnything on robotics-relevant datasets, providing insights into their practical

applicability for mobile robotics applications. The work is expected to produce publication-quality results suitable for submission to a computer vision or robotics conference. There will also be opportunities to contribute to the open-source implementations of these methods.

Required Skills

The candidate should have a strong background in computer vision and 3D geometry, with solid programming skills in Python. Familiarity with deep learning frameworks like PyTorch and knowledge of classical methods such as Structure-from-Motion and camera calibration will be valuable. Experience with Linux environments and collaborative development tools is expected, along with strong motivation for independent research work.

Practical Information

This is a 5-6 month Master 2 level internship based at Institut Pascal, Université Clermont Auvergne. The position offers access to GPU computing resources and robotic platforms. Depending on the results and mutual interest, there is a possibility of continuation as a PhD position.

Contact

Prof. Thierry Chateau
Full Professor
Institut Pascal, Université Clermont Auvergne
Email: thierry.chateau@uca.fr

References

[1] Jianyuan Wang et al., "VGGT: Visual Geometry Grounded Transformer", arXiv:2503.11651v1, 2025.

[2] Nikhil Keetha et al., "MapAnything: Universal Feed-Forward Metric 3D Reconstruction", arXiv:2509.13414v2, 2025.