Hybrid (solved-learned) Camera Localization

Abdul Rahman Samra(ИТМО)

Scientific supervisor – Candidate of Technical Sciences, Professor. Sergey V. Shavetov (ИТМО)

Introduction. Camera localization is a crucial problem in computer vision and robotics, with applications ranging from autonomous navigation to augmented reality. Traditional localization techniques, such as Structure from Motion (SfM) and Simultaneous Localization and Mapping (SLAM), rely on geometric and feature-based methods. While these approaches offer high accuracy in structured environments, they struggle with challenges like dynamic scenes, occlusions, and textureless areas. On the other hand, deep learning-based localization methods leverage neural networks to learn complex scene representations but often suffer from high computational costs, generalization issues, and dependency on large-scale labeled datasets.[1,2]

This research proposes a hybrid approach that integrates classical geometric techniques with neural network-based models to achieve robust and efficient camera localization. The goal is to develop a model that overcomes the limitations of both methods, enhancing robustness against noise, occlusions, and environmental variations. By combining the interpretability of classical methods with the adaptability of deep learning, this approach aims to improve localization accuracy and reliability, making it suitable for real-world applications in urban environments.[3]

Main part.

1. Background and Problem Statement: This section discusses the fundamentals of camera localization, its significance in applications like robotics and augmented reality, and the limitations of existing methods. It highlights the challenges faced by classical and neural network-based approaches, such as sensitivity to environmental conditions and data dependency.

2. Methodology and Framework: Here, we detail the proposed hybrid approach that integrates classical geometric techniques (e.g., SfM, SLAM) with neural networks. The section explains how traditional methods ensure geometric consistency while neural networks enhance robustness and adaptability.

3. Performance Evaluation and Benchmarking: This part presents experimental setups, datasets, and evaluation metrics used to compare the hybrid model against conventional methods. Results focus on accuracy, robustness, and noise immunity.

Conclusions. This research demonstrates that a hybrid approach enhances camera localization by combining classical and neural methods, improving accuracy, robustness, and noise immunity for real-world applications.

References

1. Rockwell C., Johnson J., Fouhey D.F. The 8-Point Algorithm as an Inductive Bias for Relative Pose Prediction by ViTs. 2022.
2. Song X. et al. TransBoNet: Learning camera localization with Transformer Bottleneck and Attention // Pattern Recognit. Elsevier Ltd, 2024. Vol. 146.
3. Rockwell C. et al. FAR: Flexible, Accurate and Robust 6DoF Relative Camera Pose Estimation. 2024.

Author: _______________ Samra Abdul Rahman

Scientific supervisor: ____________________ Shavetov Sergey Vasilievich