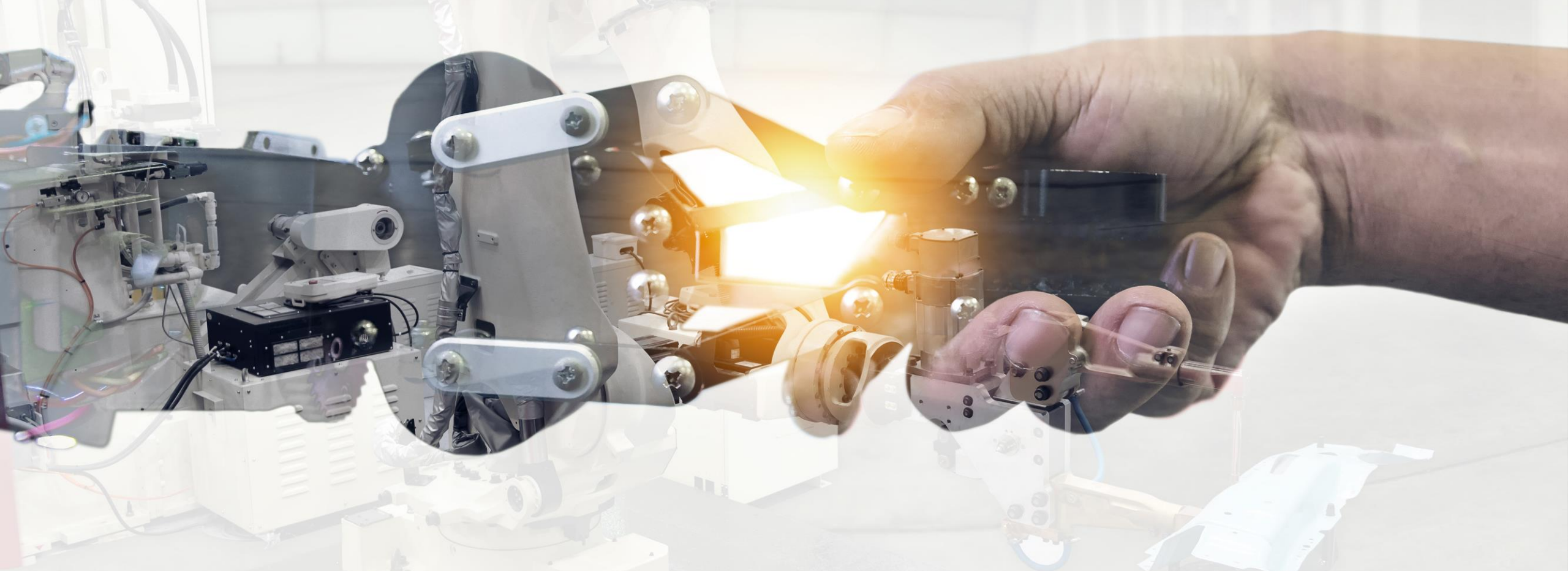


Collaborative visual-inertial localization of teams with floorplan extraction

Sándor Gazdag, Dániel Pásztornicky, Zsolt Jankó, Tamás Szirányi, András Majdik

2026. 01. 08.

ARTIFICIAL INTELLIGENCE National Laboratory
HUN REN | SZTAKI

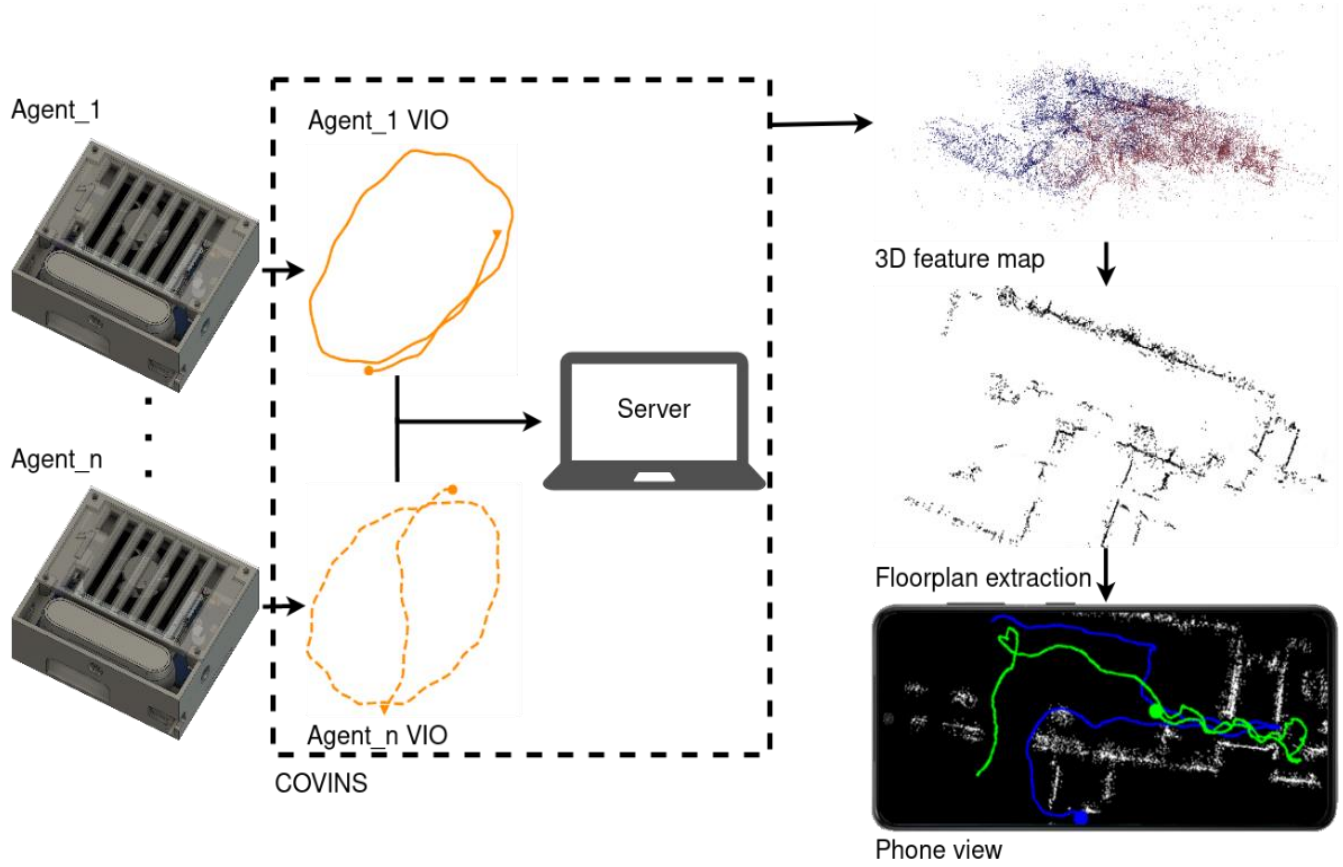
Motivation and contribution

Motivation:

- GNSS is unavailable or have large error inside
- Sparse feature maps are hard to interpret for humans

Contribution:

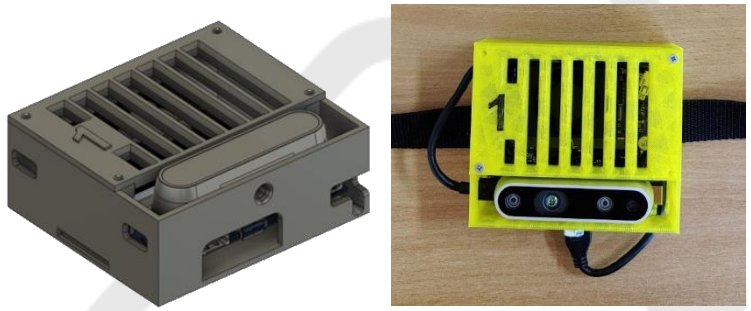
- Human-interpretable floorplan-like map in real-time
- Compact wearable hardware prototype for human or robotic agents



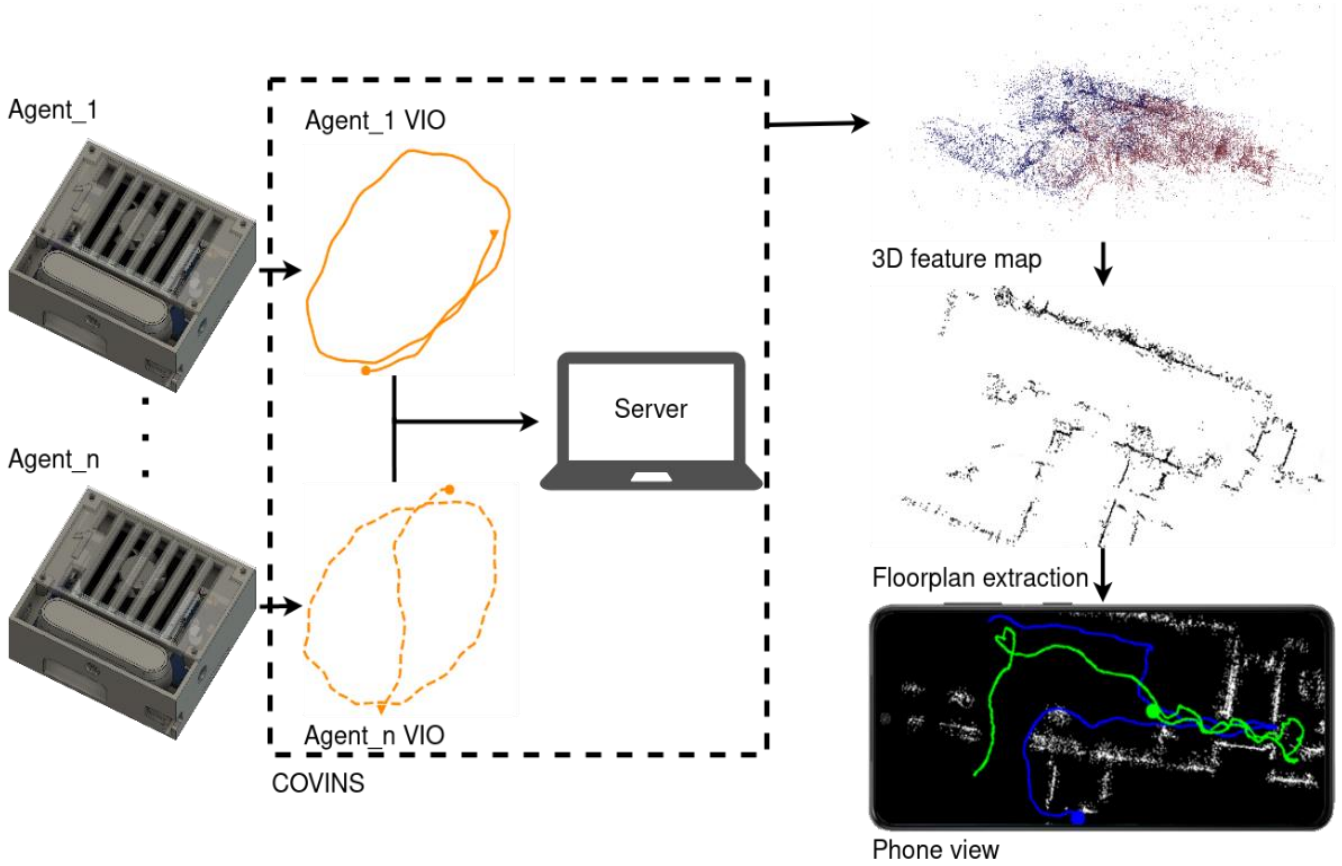
Flowchart of the algorithm

Architektúra

- Agents:
  - Local visual inertial odometry [1]
  - NVIDIA Jetson Xavier NX
- Server:
  - Centralized C-SLAM [2]
  - Floorplan extraction
- ROS + WiFi + Webserver + Phones



3D model and assembled hardware



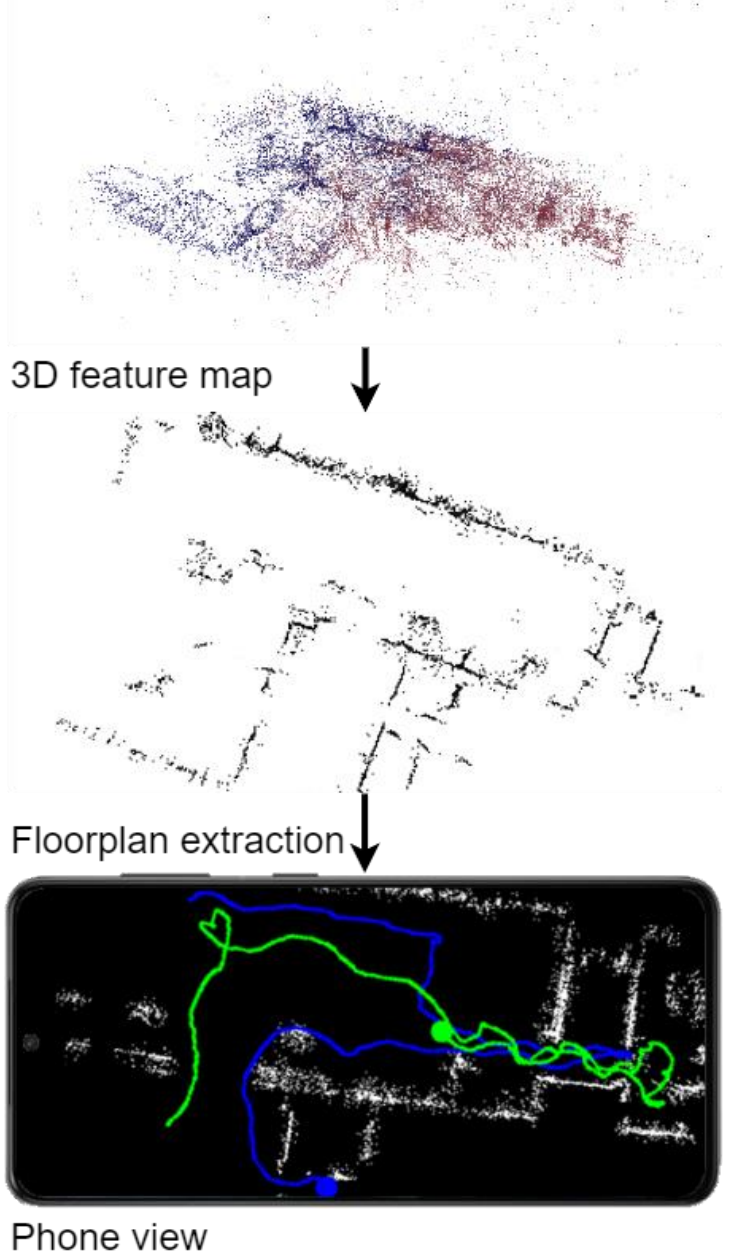
Flowchart of the algorithm

[1] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel and Juan D. Tardós, „ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM", *in IEEE Transactions on Robotics , 2021.*

[2] P. Schmuck, T. Ziegler, M. Karrer, J. Perraudin, and M. Chli, „Covins: Visual-inertial slam for centralized collaboration*", in proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), 2021.*

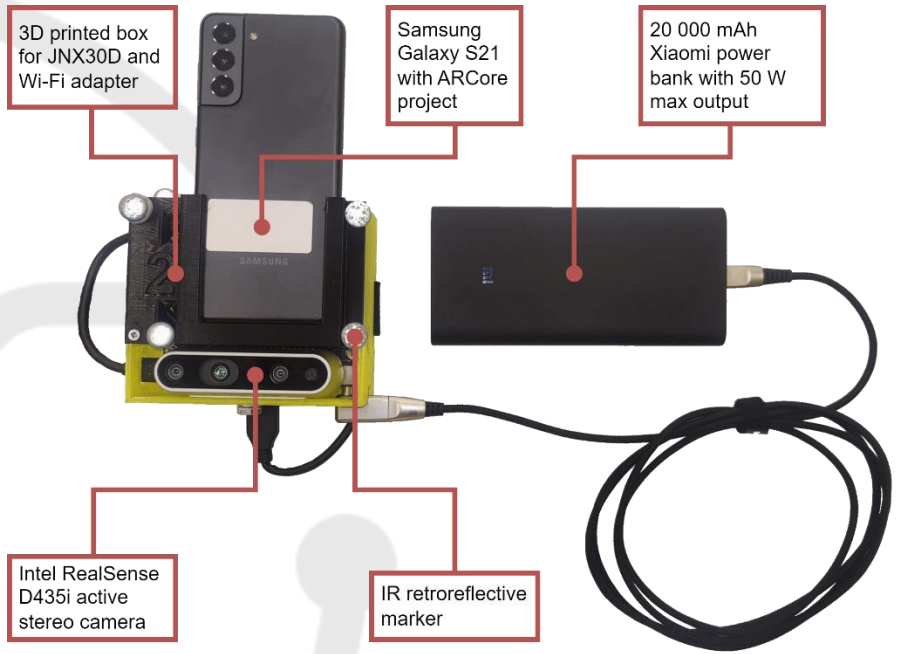
Real-time floorplan extraction

- Sparse feature cloud on server
- Statistical Outlier Removal
- Vertical histogram for floor and ceiling removal
  - Search for local maxima
  - Remove planes
- Orthographic projection
- Plotting trajectories
- Publish to webserver
- View on mobile

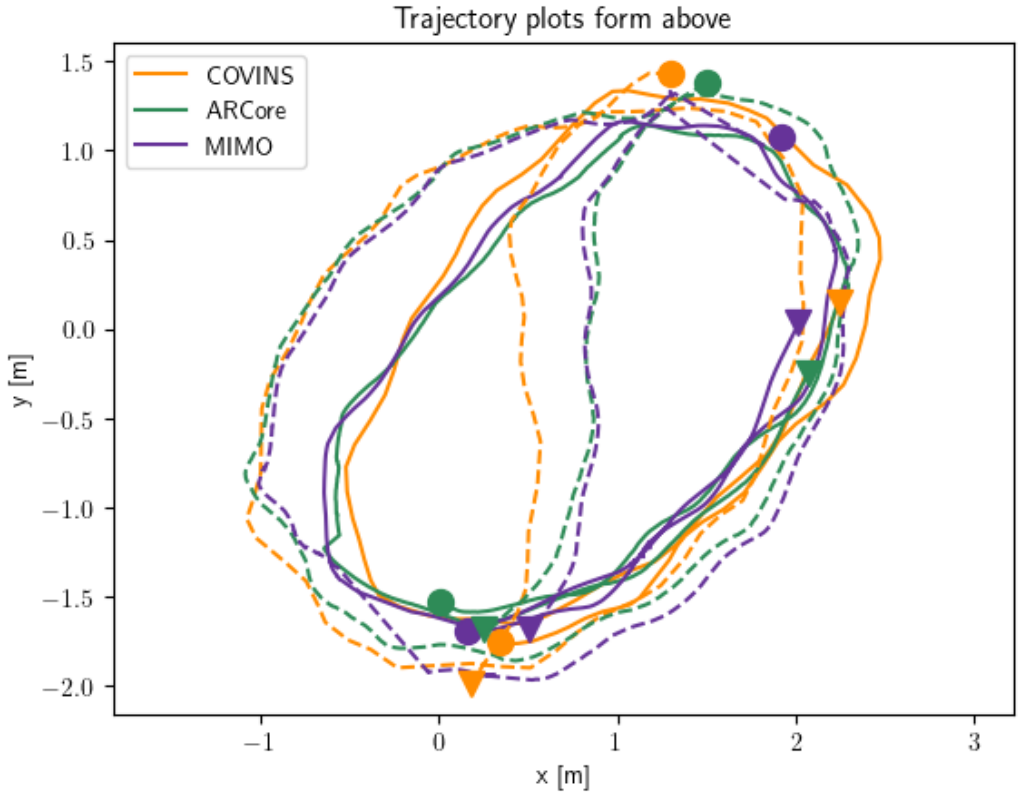


Results

- GT position from motion capture arena (sub mm accuracy)
- Similar accuracy to a solution based on Google ARCore
- Tested in real environment (video)
- Real-time



Test hardware



Comparison of trajectories



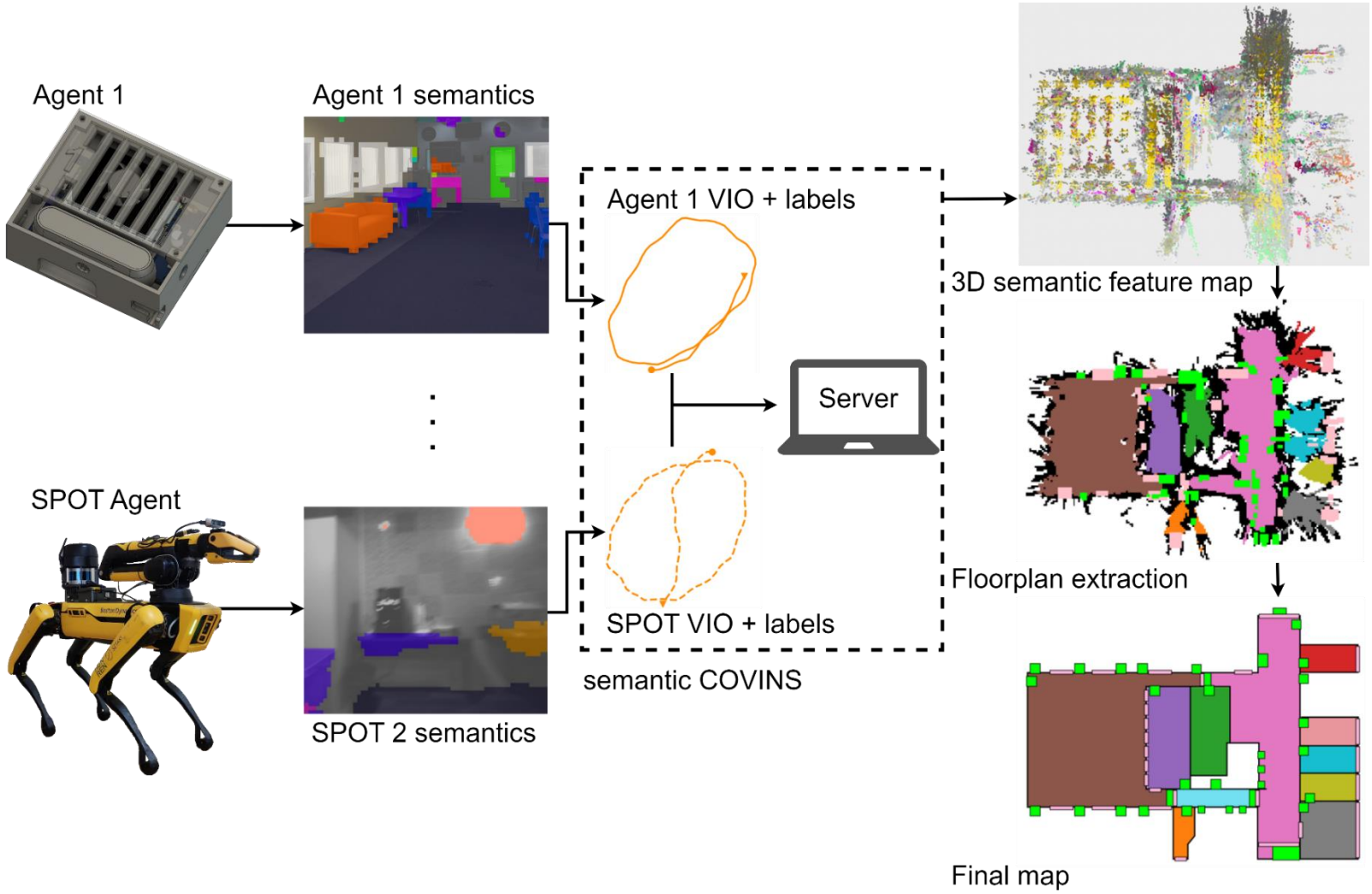
Current Research directions

SPOT as agent

- More sensor data
- Automated patrol

Semantic information

- Filtering features
- High level map
- Semantic loop closure



Sparse semantic map

Client 1

Client 2

No Image

Client 3

Thank you for the attention!

sztaki.hun-ren.hu

Acknowledgements

This work was supported by the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory, by the Hungarian Scientific Research Fund (No. NKFIH OTKA KH-126513), and by the Force Modernization and Transformation Command of the Hungarian Defence Forces. The research reported in this paper is part of project no. TKP2021-NVA-01, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme. The project was also supported by the DCEP program, funded by the NRDI Fund of the Ministry of Culture and Innovation and BME, under a grant agreement with the NRDI Office.

References

[1] X. Liu, H. Peng, N. Zheng, Y. Yang, H. Hu, and Y. Yuan, „Efficientvit: Memory efficient vision transformer with cascaded group attention”, *in proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.*

[2] S. Gazdag, D. Pasztornicky, Z. Jankó, T. Szirányi and A. L. Majdik, „Collaborative Visual-Inertial Localization of Teams With Floorplan Extraction”, *in proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), 2023*

[3] P. Schmuck, T. Ziegler, M. Karrer, J. Perraudin, and M. Chli, „Covins: Visual-inertial slam for centralized collaboration*”, in proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), 2021.*

[4] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel and Juan D. Tardós, „ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM”, *in IEEE Transactions on Robotics , 2021.*

[5] Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso and Antonio Torralba, „ Semantic Understanding of Scenes through ADE20K Dataset”, *in International Journal on Computer Vision (IJCV)*