Enhancing Visual SLAM for Autonomous Forest Navigation through Robust Feature Correspondence

Department of Computing Imperial College London

Author:
Saifullah Ijaz

Supervisors:
Dr Bahadir Kocer
Dr Ronald Clark

Forestry Applications

- Approximately 31% of the Earth's land mass is made up of forests.
- Typical forestry tasks include:
- Plant and wildlife monitoring
- Early detection of wildfires
- Search and rescue operations

Showalter, G. (2024) International Day of forests 2024, British Ecological Society. Available at: https://www.britishecologicalsociety.org/international-day-of-forests-2024/#:~:text=Forests%20make%20up%2031%25%20of,vital%20ecosystems%20on%20the%20planet.

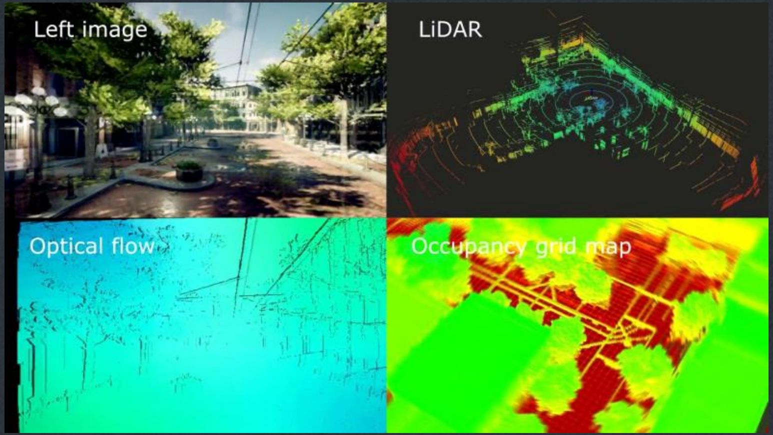
Using Robots in Forests

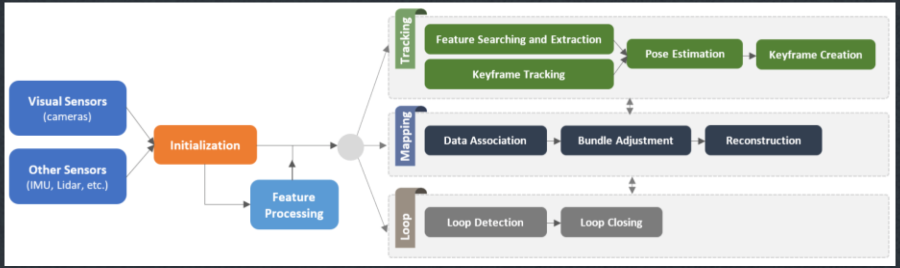
- Increase efficiency
- Reduce safety risk
- Meaningful data insights
- However, forests present unique navigation challenges
- GPS denied environments

Lopez, I. (2020) Treeswift's autonomous robots take flight to save forests, Penn Engineering Blog. Available at: https://blog.seas.upenn.edu/treeswifts-autonomous-robots-take-flight-to-save-forests/
Bastos, A.S. and Hasegawa, H. (2013) 'Behavior of GPS signal interruption probability under tree canopies in different forest conditions', European Journal of Remote Sensing, 46(1), pp. 613–622. doi:10.5721/eujrs20134636.

SLAM

- Robot moves through the environment
- Re-observed features are used for localisation and mapping tasks
- Localisation -> Where am I?
- Mapping -> Where are the objects?

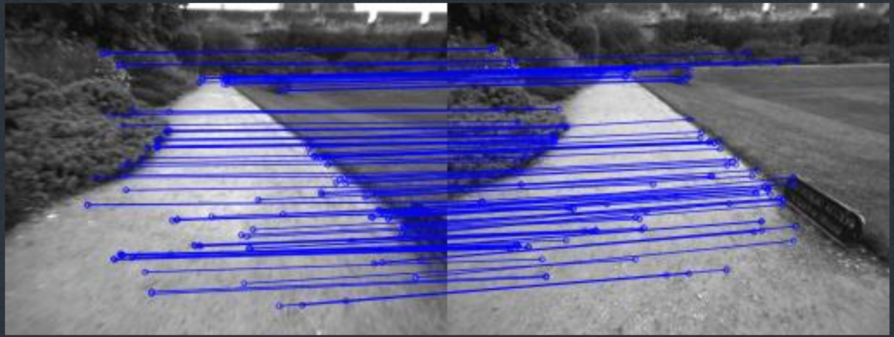




Wang, W. et al. (2020) Tartanair: A dataset to push the limits of visual slam, arXiv.org. Available at: https://arxiv.org/abs/2003.14338
Tourani, A. et al. (2022) Visual slam: What are the current trends and what to expect?, arXiv.org. Available at: https://arxiv.org/abs/2210.10491

Feature detection and correspondence

- Feature detection -> what points in the image are easily recognisable?
- Feature correspondence -> where are the same features in two different frames?

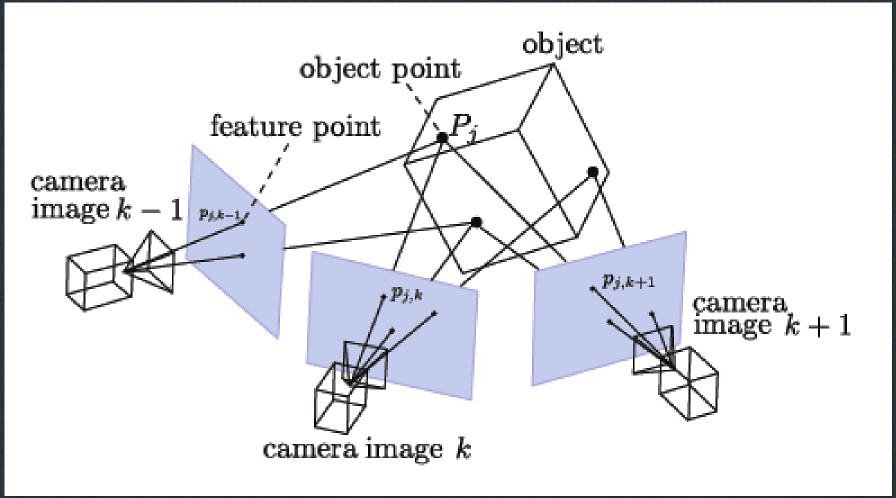


ORB feature matching

Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Transactions on Robotics, 31(5), 1147-1163.

Pose Estimation

- As the robot moves through the environment
- Features are re-observed between frames
- Translation and rotation can be estimated using the intrinsic parameters of the camera
- Relative poses are combined to form an overall trajectory



Comet Labs Research Team (2017) Teaching robots presence: What you need to know about slam, Medium. Available at: https://blog.cometlabs.io/teaching-robots-presence-what-you-need-to-know-about-slam-9bf0ca037553

Challenges with Forest Environments

01 **Swaying tree branches due to wind**

-> Motion blur

02 **Dappled sunlight through tree canopy**

-> Intermittent lighting conditions

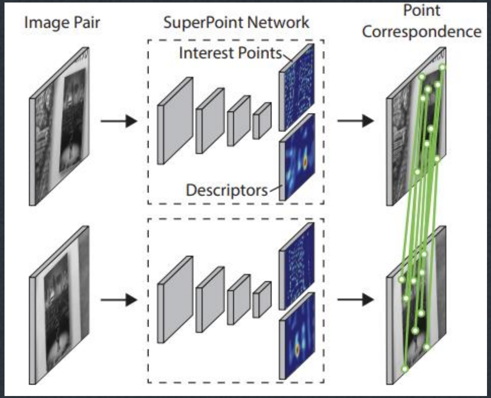
03 **Grass or leaves covering forest floor**

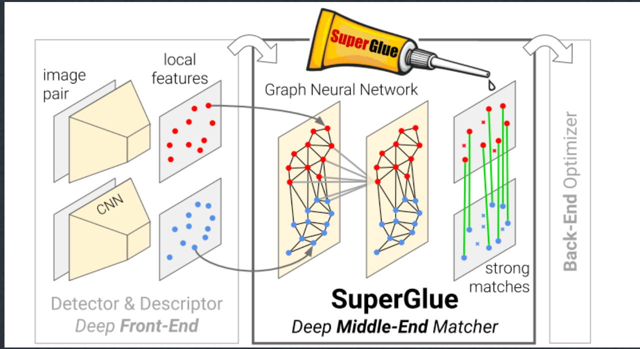
-> Monotononous textures

Wind+Blowing+Trees images – browse 77,447 stock photos, vectors, and video Adobe Stock. Available at: https://stock.adobe.com/uk/search?k=wind%2Bblowing%2Btrees
Adam (2019) Komorebi: When sunlight shines through trees: Awa Tree Blog., AWA Trees. Available at: https://www.awatrees.com/2017/02/16/komorebi-sunshine-through-trees/
Panupattanagul, K. AI generated forest floor was covered in a thick layer of fallen leaves and debris., Vecteezy. Available at: https://www.vecteezy.com/photo/25247886-ai-generated-forest-floor-was-covered-in-a-thick-layer-of-fallen-leaves-and-debris

SuperPoint and SuperGlue

- Traditional feature detectors and matchers fail under the challenging forest conditions
- Deep learning models have been trained as robust alternatives
- SuperPoint -> Interest Point Detector
- SuperGlue -> Feature Matcher

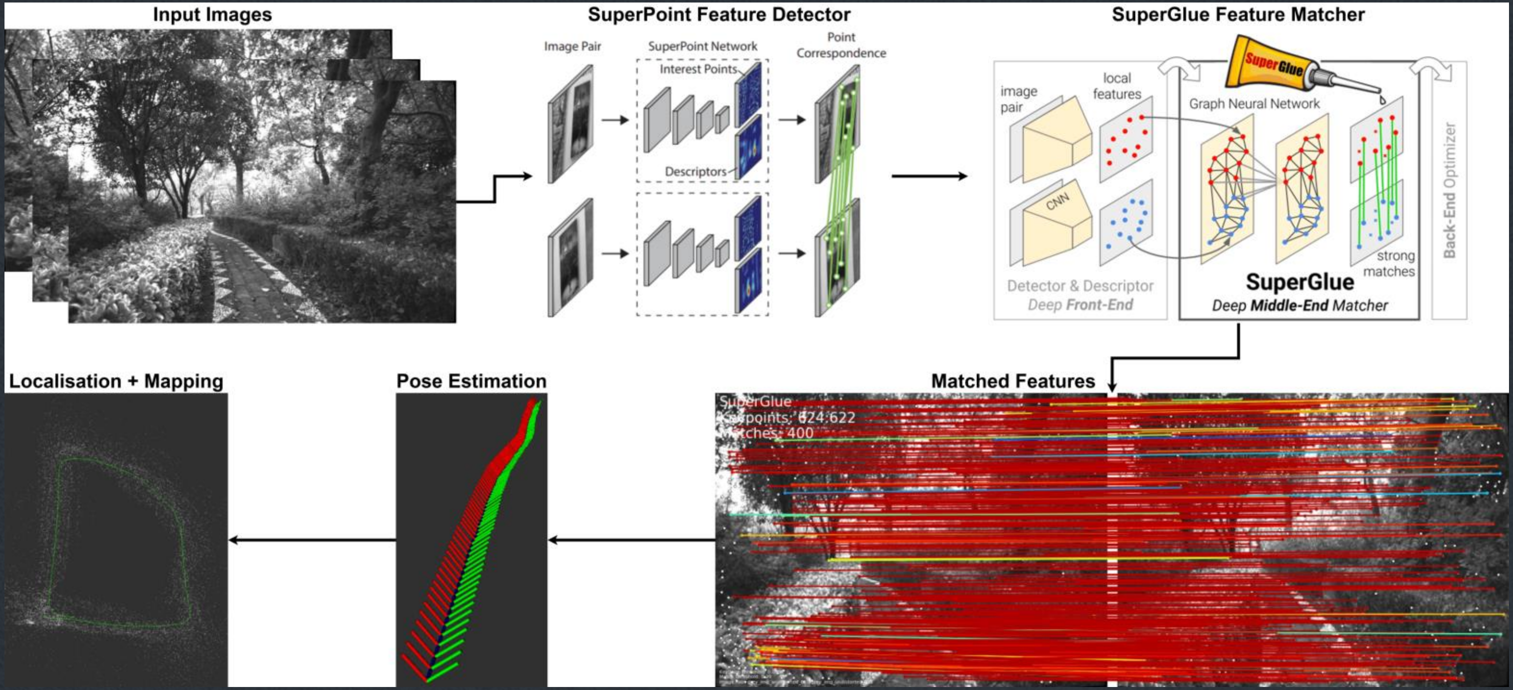




DeTone, D., Malisiewicz, T. and Rabinovich, A. (2018) SuperPoint: Self-supervised interest point detection and description, arXiv.org. Available at: https://arxiv.org/abs/1712.07629
Sarlin, P.-E. et al. (2020) Superglue: Learning feature matching with Graph Neural Networks, arXiv.org. Available at: https://arxiv.org/abs/1911.11763

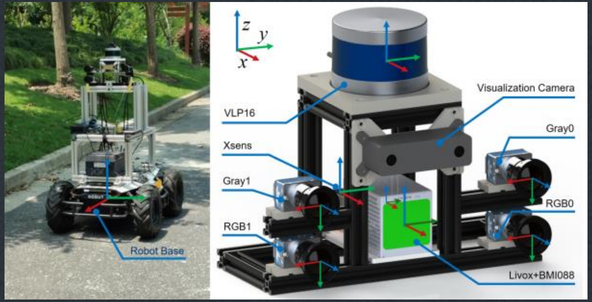
How effective are SuperPoint and SuperGlue for robust localisation when incorporated as part of a SLAM system for forest environments?

Forest SLAM


Input Images
SuperPoint Feature Detector
Image Pair
SuperPoint Network
Interest Points
Descriptors
Point Correspondence
SuperGlue Feature Matcher
image pair
local features
CNN
Detector & Descriptor
Deep Front-End
Graph Neural Network
SuperGlue
Deep Middle-End Matcher
strong matches
Back-End Optimizer
Localisation + Mapping
Pose Estimation
Matched Features
SuperGlue
Matches: 400

BotanicGarden Dataset

- Liu et al. released the BotanicGarden dataset in 2024
- Contains unstructured natural scenes with a mixture of thick woods, narrow trails and grasslands
- Highly accurate ground truth pose data
- Leaderboard available comparing current state-of-the-art SLAM systems



Liu, Y. et al. (2024) Botanicgarden: A high-quality dataset for robot navigation in unstructured natural environments, IEEE Xplore. Available at: https://ieeexplore.ieee.org/document/10415477

Evaluation

BotanicGarden Leaderboard

- Displays average Relative Pose Error (RPE) across seven sample sequences

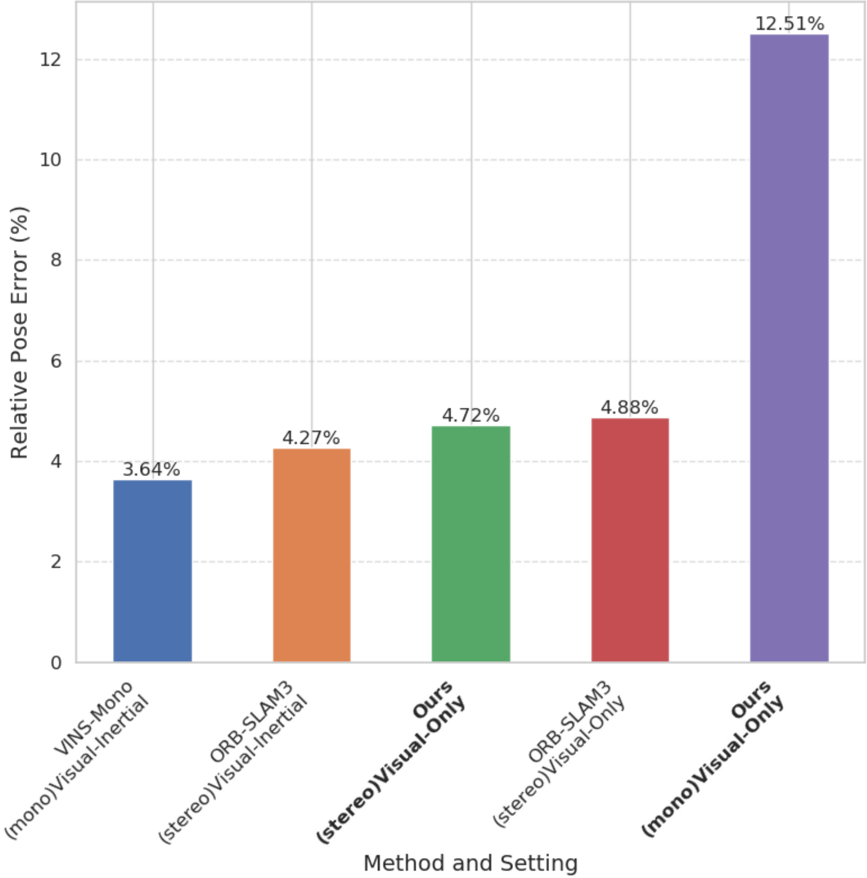
- RPE is the difference between the estimated relative pose and the ground truth relative pose

- Calculated over fixed-length segments

- Unaffected by the trajectory length so it is invariant to odometry drift

Liu, Y. et al. (2024) Botanicgarden: A high-quality dataset for robot navigation in unstructured natural environments, IEEE Xplore. Available at: https://ieeexplore.ieee.org/document/10415477

BotanicGarden leaderboard showing the RPE for different SLAM systems averaged across seven sample sequences expressed as a percentage.



Ablation Study

- Artificially add motion blur to images to create challenging scenarios

- Motion blur kernel averages pixel intensities across a particular direction simulating motion blur

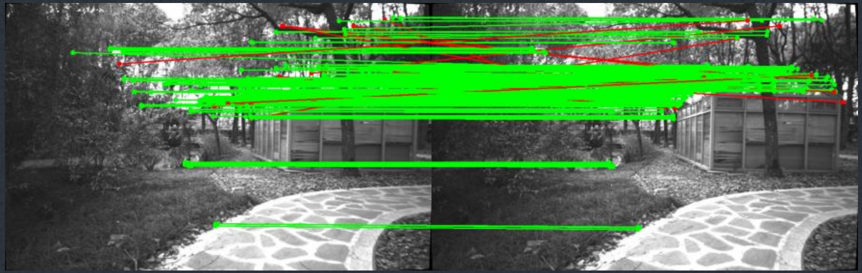
- Substitute SuperPoint + SuperGlue with ORB + BF to see which combination is the most robust

- Compare pose estimations when incorporating both into Forest SLAM

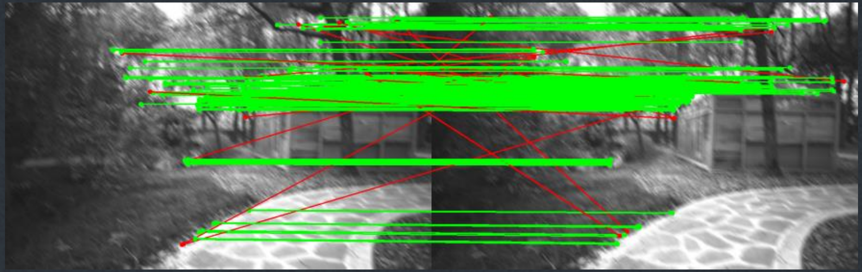
Feature Matching OpenCV. Available at: https://docs.opencv.org/4.x/dc/dc3/tutorial_py_matcher.html

ORB + BF

- ORB features are mainly found on easy to distinguish points like tree trunks and corners of objects
- Features are concentrated in central parts of the frame
- Even with no additional motion blur there are several incorrect matches
- Additional motion blur increases number of incorrect matches significantly



ORB feature matching with no additional motion blur

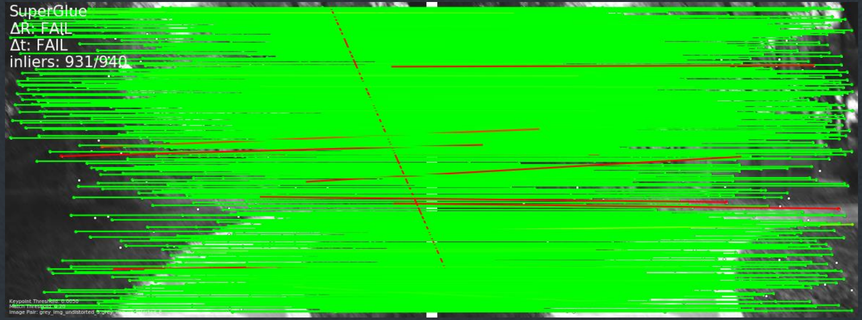


ORB feature matching with a motion blur kernel size of 10

SuperPoint + SuperGlue

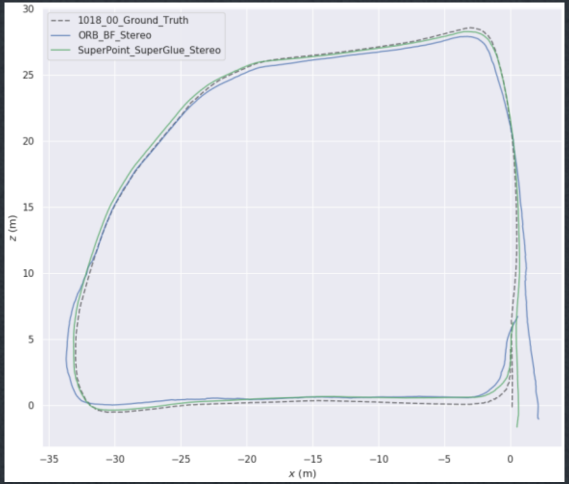
- SuperPoint features are spread out across the entire frame
- With no additional motion blur, SuperGlue correctly matches all features
- Even with additional motion blur there are only a few incorrect matches
- The ratio of correct to incorrect matches remains very high

SuperGlue feature matching with no additional motion blur

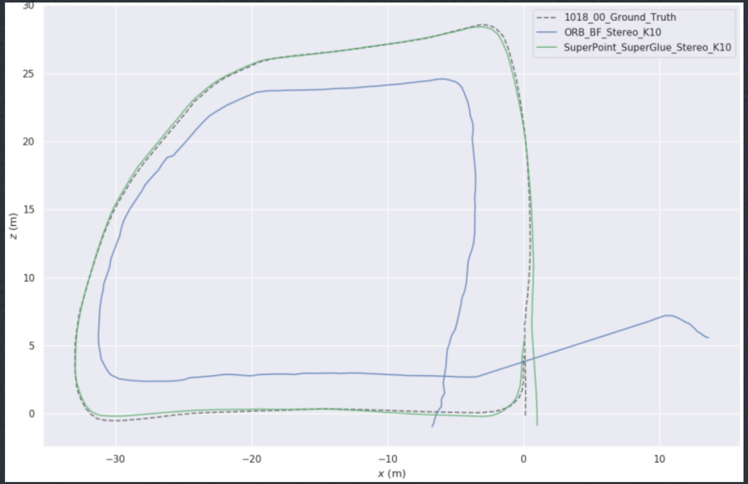


SuperGlue feature matching with a motion blur kernel size of 10

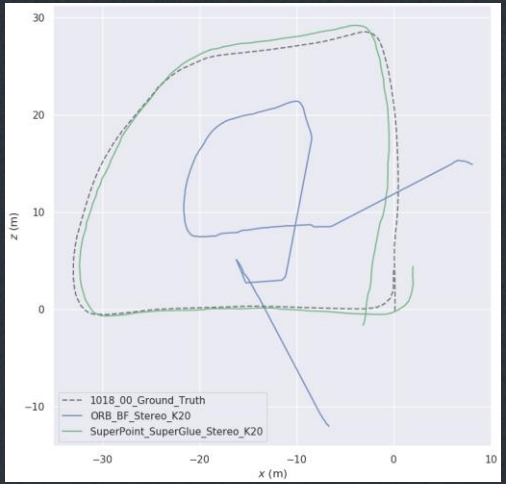
Forest SLAM Comparisons



No additional motion blur



Motion blur kernel size = 10



Motion blur kernel size = 20

- Additional motion blur causes incorrect ORB feature matching, resulting in incorrect pose estimations
- Learning-based approaches are robust, estimated poses still closely match the ground truth

Conclusions

Robust Feature Correspondence

Visual SLAM requires robust feature correspondence for accurate relative pose estimations

Traditional Approaches

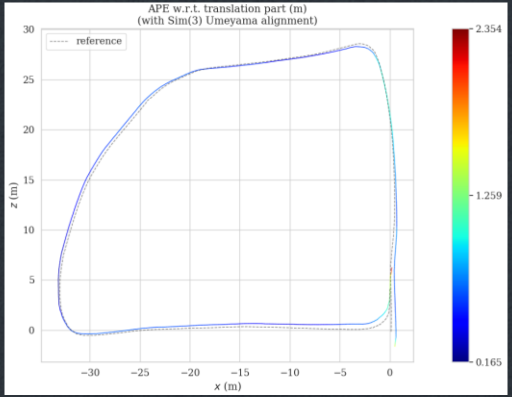
Hand-crafted feature techniques struggle to deal with dynamic forest scenes

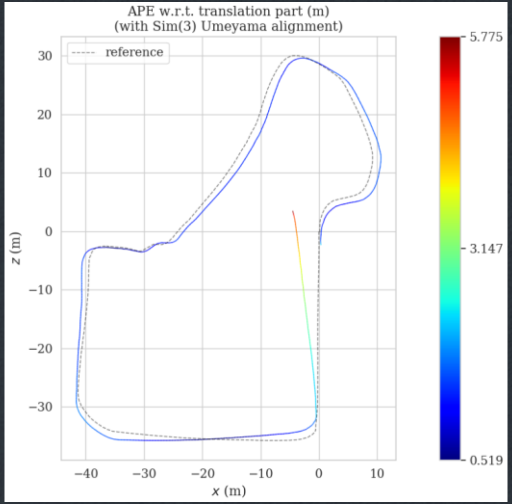
Learning-based Approaches

SuperPoint and SuperGlue are a robust combination for use in a SLAM system for forest environments

Future Work

- Bundle adjustment and loop closure detection to minimise drift
- Incorporate additional sensor modalities
- Re-train SuperGlue or DynamicGlue on forest data
- Train a deep learning model to perform pose estimations using the matched keypoints





Huber, T., Schaefer, S. and Leutenegger, S. (2024) Dynamicglue: Epipolar and time-informed data association in dynamic environments using graph neural networks, arXiv.org. Available at: https://arxiv.org/abs/2403.11370

Thank you for listening!

Questions?