Keypoint Semantic Integration for Improved Feature Matching in Outdoor Agricultural Environments

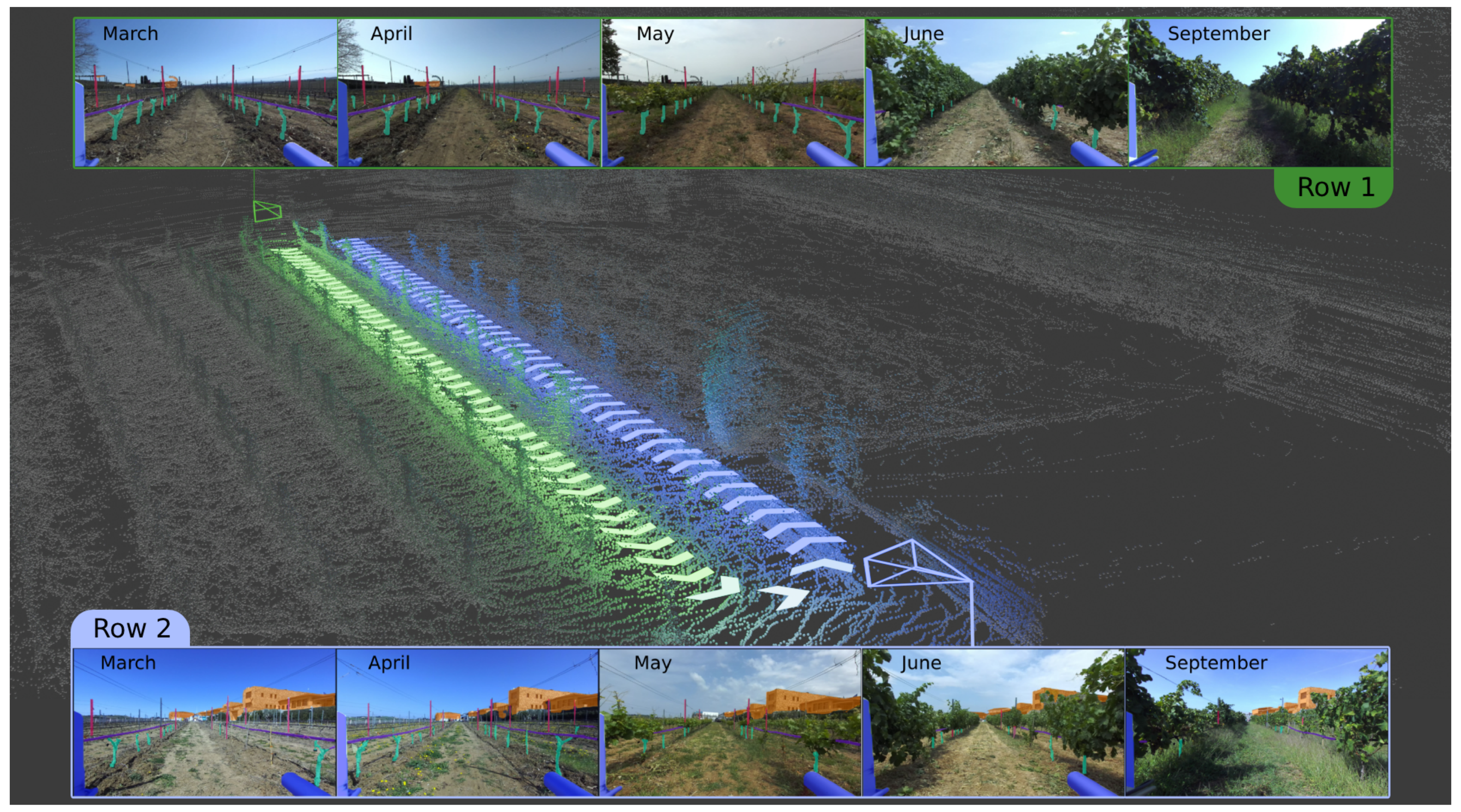
Rajitha de Silva[1] · Jacob Swindell[1] · Jonathan Cox[1] · Marija Popović[2] · Cesar Cadena[3] · Cyrill Stachniss[4] · Riccardo Polvara[1]

1. Lincoln Centre for Autonomous Systems (L-CAS), University of Lincoln, U.K. | 2. MAVLab, TU Delft, Netherlands.
3. Robotics Systems Lab, ETH Zurich, Switzerland. | 4. University of Bonn, Germany.

ICRA IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION 2026 VIENNA

Motivation

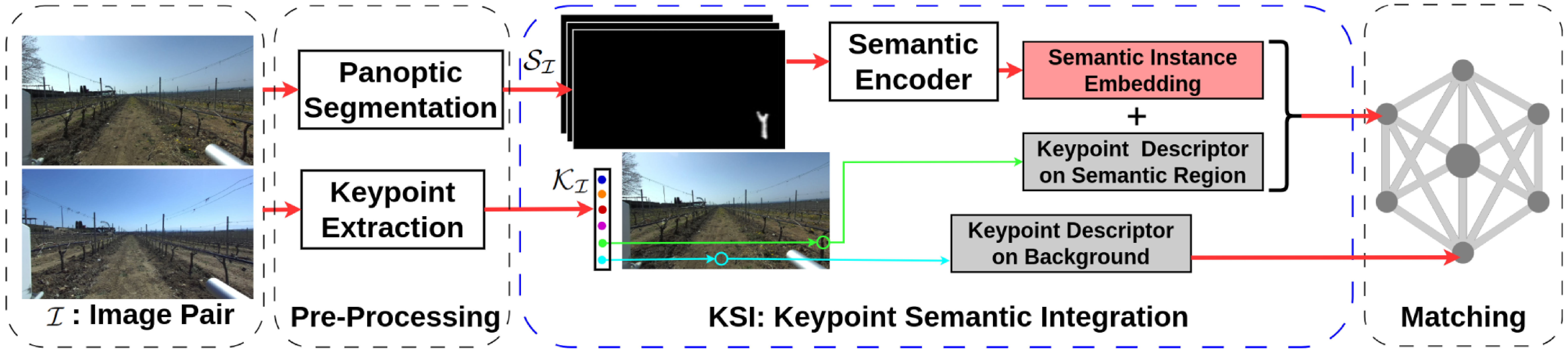
Understanding how semantic context can reduce descriptor ambiguity is key to improving visual correspondence in repetitive, changing natural environments.



Method

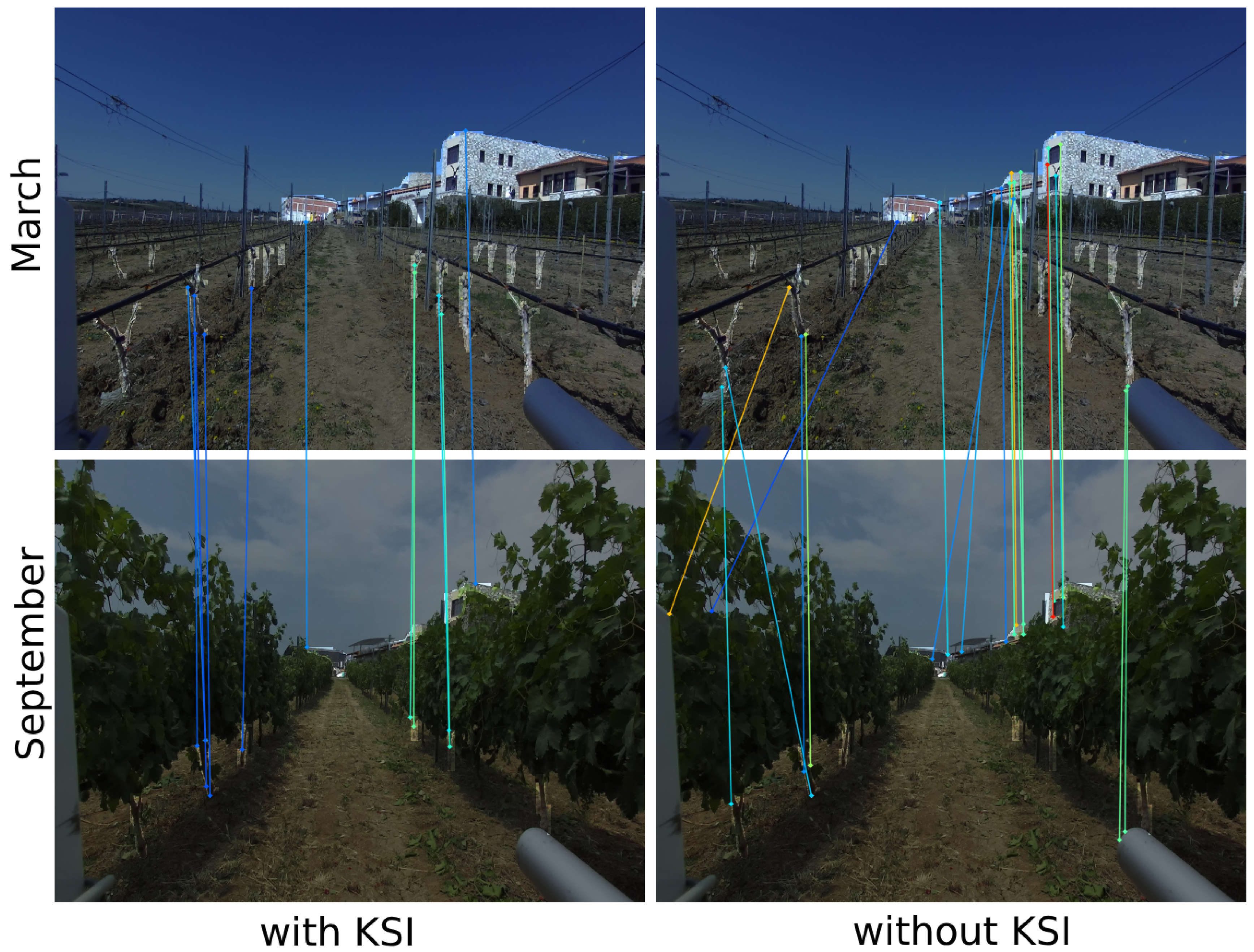
KSI operates as a plug-and-play enhancement over existing keypoint pipelines:

1. Panoptic Segmentation – A segmentation model extracts vineyard-relevant objects from RGB images.
2. Semantic Encoding – Instance masks are encoded into compact semantic embeddings.
3. Descriptor Fusion – Semantic embeddings are fused with keypoint descriptors and L2-normalised.
4. Matching – Enhanced descriptors are matched using SuperGlue [1], LightGlue [2], or similar matchers.



Results

- Keypoint matching accuracy: KSI improves matching accuracy by up to 18.87% in semantically significant vineyard regions.
- Relative pose estimation: KSI reduces relative pose error by up to 70.92%, improving trajectory estimation.
- Visual localisation: KSI improves cross-season visual localisation by up to 35.06% in challenging vineyard rows.
- Generalisation: KSI transfers beyond vineyards, improving woodland keypoint matching by 10.16% and reaching 94.66% accuracy.
- Frame rate: The system runs at 14.75 image pairs/s on desktop and 4.71 image pairs/s on Jetson AGX Orin.



Conclusion

KSI shows that adding semantic context to keypoint descriptors reduces perceptual aliasing and improves matching, pose estimation, and localisation in challenging outdoor environments. Its modular design keeps it compatible with existing visual SLAM pipelines while generalising beyond vineyards.

Limitation: KSI currently encodes instance shape and position rather than explicit class-level meaning, so performance depends on choosing semantic regions that are visually distinctive and useful.

Future work: Incorporating richer class-aware semantics could make descriptor enhancement more meaningful and reduce sensitivity to manual class selection.

References

1. Sarlin, Paul-Edouard, et al. "Superglue: Learning feature matching with graph neural networks." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
2. Lindenberger, Philipp, Paul-Edouard Sarlin, and Marc Pollefeys. "Lightglue: Local feature matching at light speed." Proceedings of the IEEE/CVF international conference on computer vision. 2023.