PV-SLAM: Panoptic Visual SLAM with Loop Closure and Online Bundle Adjustment

Ashok Bandyopadhyay[1,2], Pranjal Baranwal[2], Arijit Sur[2], Rajeev UP[1]

[1]Vikram Sarabhai Space Centre, Indian Space Research Organization, Thiruvananthapuram, India

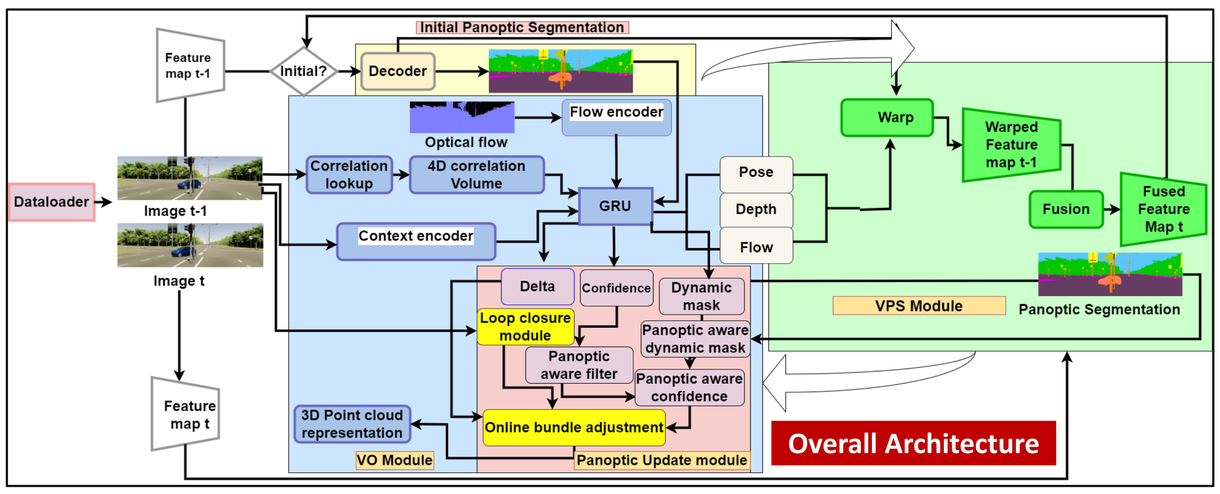
[2]Department of CSE, Indian Institute of Technology Guwahati, India

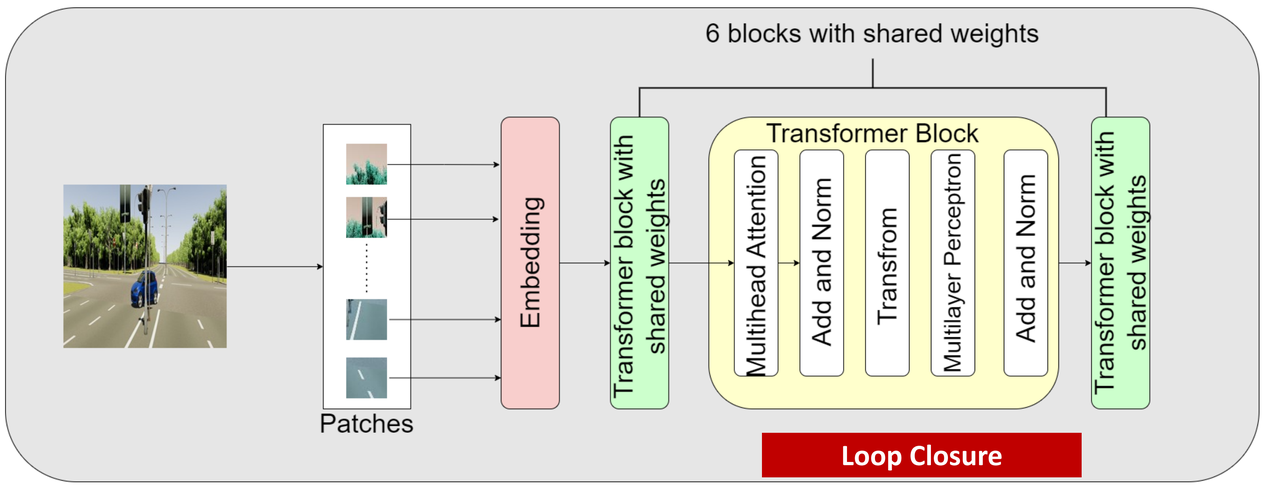
Paper # 936

Motivation

- ✓ Visual SLAM (vSLAM) uses camera data to map an environment and track a robot's position simultaneously.
- ✓ Panoptic segmentation-based Visual SLAM offers a more comprehensive, efficient, and robust solution for robot perception.
- ✓ Recent panoptic Visual SLAM methods overlook loop closure and rely on offline bundle adjustment, which can cause drift errors, emphasizing the need for online bundle adjustment.
- ✓ We introduce PV-SLAM, a novel architecture that integrates loop closure and online bundle adjustment, expanding on the PVO model.

Proposed Architecture





- ✓ Initial Panoptic Segmentation module takes an image as input and produces panoptic segmentation, which merges the semantic segmentation and instance segmentation of the image. The output result is fed into the VO module and the VPS module.
- ✓ The VO and VPS modules further improve the panoptic segmentation results

Bundle Adjustment:

- ✓ We use an online version of bundle adjustment, which applies bundle adjustment as soon as a new frame is encountered.

Loop Closure:

- ✓ The loop closure module starts by first patching the image into several parts. Then, each patch x^k is fed into the vision transformer block.
- ✓ The vision transformer then creates a feature vector out of it. Then, we can recognize loop closure when two frames have a similarity score higher than a preset threshold.

Experiments

Dataset: KITTI | Virtual KITTI | EuRoC

Training: Pre-trained image transformer named DEIT(Data Efficient Image Transformers)(base distilled variant) is used for loop closure detection.

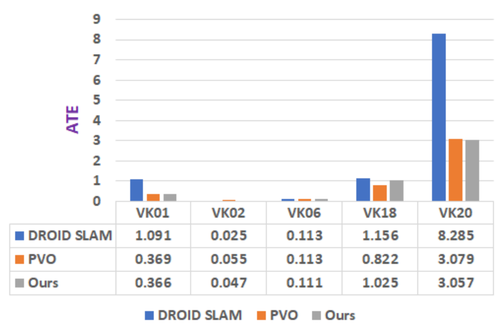
Working Principle

Algorithm 1 Complete algorithm for our Panoptic Segmentation based Visual SLAM

Require: Frame: Images acquisition
/*Initial Panoptic Segmentation*/
while Frame I_i **do**
  $P_i \leftarrow PanopticFPN(I_i)$
end while
/*VO module*/
while Frame I_i **do**
  $g_{\theta i} \leftarrow FeatureVector$; $c_\theta \leftarrow ContextVector$; $Pose, Flow, Depth \leftarrow Backbone(I_i)$;
  while j such that i, j share common areas **do**
    $C_{ij} \leftarrow \langle g_\theta I_i, g_\theta I_j \rangle$
  end while
  $F_i \leftarrow Flow$; $Delta, Confidence, DynamicMask \leftarrow GRU(C_i, F_i, P_i)$;
  $PanopticAwareDynamicMask \leftarrow Sigmoid(DynamicMask, P_i)$
  if LoopDetection(i) = True **then**
    Add neighbour edges in the frame graph i,j-1
  end if
  $BundleAdjustment(I_i, PanopticAwareConfidence)$; $Genferate3DPointCloud(I_i)$;
end while
/*VPS module*/
while Frame t **do**
  $WarpedFeaturest \leftarrow Fusion(Depth, Flow, Pose, FeatureMapt)$
end while

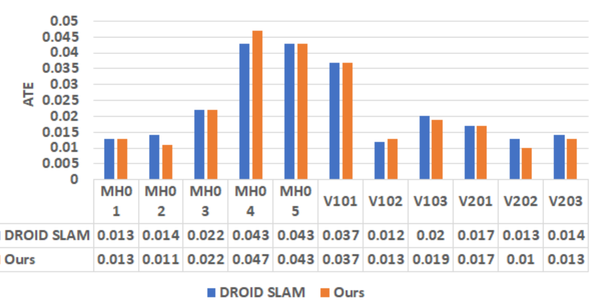
Result

Evaluation Metrics: Absolute Trajectory Error(ATE) is used as evaluation metric.



	VK01	VK02	VK06	VK18	VK20
DROID SLAM	1.091	0.025	0.113	1.156	8.285
PVO	0.369	0.055	0.113	0.822	3.079
Ours	0.366	0.047	0.111	1.025	3.057

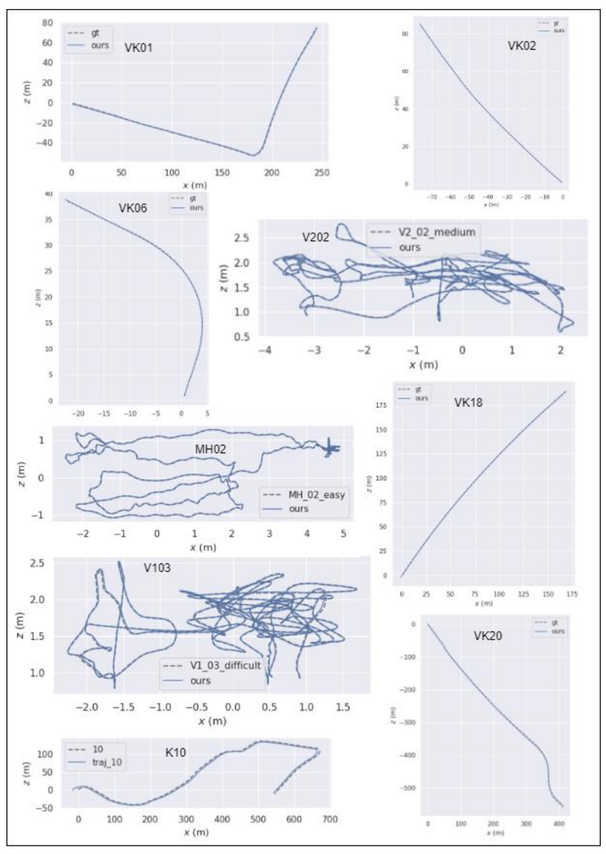
DROID SLAM PVO Ours



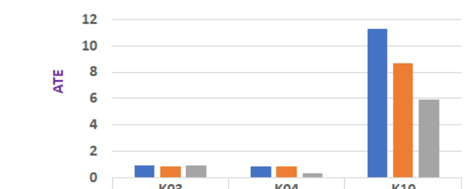
	MH0 1	MH0 2	MH0 3	MH0 4	MH0 5	V101	V102	V103	V201	V202	V203
DROID SLAM	0.013	0.014	0.022	0.043	0.043	0.037	0.012	0.02	0.017	0.013	0.014
Ours	0.013	0.011	0.022	0.047	0.043	0.037	0.013	0.019	0.017	0.01	0.013

DROID SLAM Ours

Virtual KITTI Dataset result



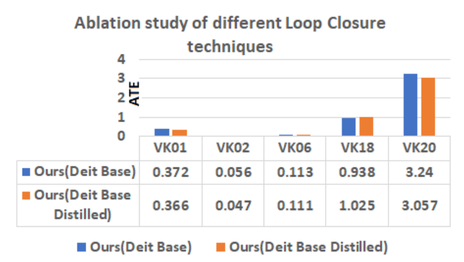
EuRoC Dataset result



	K03	K04	K10
DROID SLAM	0.893	0.816	11.308
PVO	0.855	0.808	8.66
Ours	0.916	0.333	5.882

DROID SLAM PVO Ours

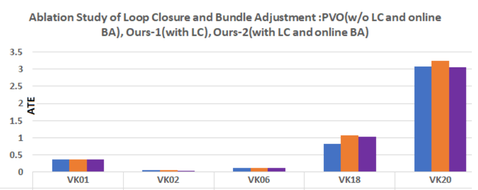
KITTI Dataset result



	VK01	VK02	VK06	VK18	VK20
Ours(Deit Base)	0.372	0.056	0.113	0.938	3.24
Ours(Deit Base Distilled)	0.366	0.047	0.111	1.025	3.057

Ours(Deit Base) Ours(Deit Base Distilled)

Ablation study: Loop closure



	VK01	VK02	VK06	VK18	VK20
PVO	0.369	0.055	0.113	0.822	3.079
Ours-1	0.372	0.049	0.112	1.07	3.238
Ours-2	0.366	0.047	0.111	1.025	3.057

PVO Ours-1 Ours-2

Generated trajectories of Kitti, Virtual Kitti and EuRoC dataset various sequences.

Ablation study: Loop closure & BA

Takeaways

Following are the key takeaways for the proposed scheme:

Comprehensive modeling of the scene: Proposed scheme enhances Visual Odometry by incorporating panoptic segmentation results, creating panoptic-aware dynamic masks for a better understanding of scene.

High Accuracy: Vision transformer-based loop closure module corrects accumulated errors. Online bundle adjustment optimizes camera trajectories in real-time.

High Adaptability: The proposed model handles dynamic objects.

High Robustness: High robustness achieved through noble Vision transformer-based loop closure detection techniques.

Reference

PVO: Panoptic Visual Odometry, CVPR 2023
DROID-SLAM: Deep Visual SLAM for Monocular, Stereo, and RGB-D Cameras, NeurIPS 2021