Switching-based Multi-modal SLAM for Extreme and Degraded Environments
(極限・劣化環境のためのマルチモーダル情報を用いた切替型 SLAM)

47-236676 Lee Junwoon (李晙源)
Department: Human and Engineered Environmental Studies
Graduated in 2024 Academic Year
Supervisor: Prof. Atsushi Yamashita

This paper presents Switch-SLAM, switching-based multi-modal SLAM for extreme and degraded environments, designed to tackle the challenges in degenerate environments for LiDAR and visual SLAM. Switch-SLAM achieves high robustness and accuracy by utilizing a switching structure that transitions from LiDAR to visual odometry when degeneration of LiDAR odometry is detected. To efficiently detect degeneration, Switch- SLAM incorporates a non-heuristic degeneracy detection method that does not require heuristic tuning and demonstrates generalizability across various environments. Switch-SLAM is evaluated on diverse datasets containing both LiDAR and visual odometry degeneracy scenarios. The experimental results highlight the accurate and robust localization by the proposed method in multiple challenging environments with either LiDAR or visual SLAM degeneracy.

Key words: SLAM, sensor fusion, localization, LiDAR degeneracy, harsh environment

1 Introduction

Simultaneous localization and mapping (SLAM) systems are subject to several limitations that arise from inherent constraints imposed by the sensors. For example, LiDAR SLAM tend to degenerate in environments lacking distinct structures. Conversely, visual SLAM face challenges in scenarios involving an aggressive motion, rapidly changing light conditions, and texture-less environments. To handle these issues, various LiDAR visual SLAM methods have been proposed, which integrate information from the LiDAR and camera. However, these methods have weaknesses when handling persistent degeneracy that exceeds the capabilities of the system. This limitation primarily arises from their reliance on fusion methods using maximum a posteriori (MAP) estimation.

To address these limitations, we propose a Switching-based LiDAR-Inertial-Visual SLAM (Switch-SLAM). Switch-SLAM[1)] parallelly processes LiDAR and visual odometry and selects the appropriate sensor odometry using non-heuristic degeneracy detection, Switch-SLAM incorporates a switching structure that effectively avoids failure information from propagating throughout the entire system, thereby mitigating the negative impact on performance. The main contributions of our work are as follows:

・Switching structure: Switching structure allows for the selection of an optimal initial guess between LiDAR and visual odometry. This selection efficiently avoids long-term degeneracy and ensures that only reliable estimations propagate through the entire system, improving overall performance.

・Non-heuristic degeneracy detection: Non-heuristic degeneracy detection checks whether the optimization process has converged or not by employing a pre-defined threshold, grounded in physical assumptions and statistical significance. This detection mechanism enhances the ability to identify degenerate situations effectively without the need for heuristic tunning of the threshold.

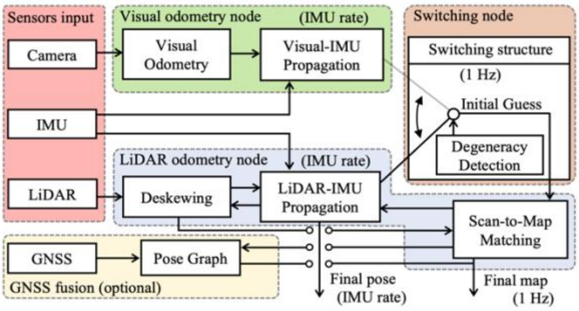

Fig. 1 The system structure of Switch-SLAM

2 Method

2.1 System Overview

The overview of the proposed method is shown in Fig. 1. The proposed approach consists of three main components: visual odometry, LiDAR odometry, and a switching node.

In the visual odometry node, the pose is estimated with sliding window optimization of tracked features. The estimated pose from visual odometry is then propagated at the frequency of the IMU measurements.

In the LiDAR odometry node, the LiDAR distortion resulting from ego-motion is corrected using the poses obtained from the switching structure. Subsequently, scan-to-map matching is conducted utilizing the geometric features, with an initial guess provided by the switching node. The estimated pose from the scan-to-map matching is also propagated at the IMU frequency.

In the switching node, the initial guess for the scan-to-map matching is selected between the poses derived from LIDAR-IMU and visual-IMU propagation, based on the reults of degeneracy detection. Our work also includes a GNSS option, which is fused with the final pose from the scan-to-map matching.

2.2 Lidar-Inertial-Visual Slam

1) LiDAR Odometry: LiDAR odometry is performed by scan-to-map matching. In this process, planar and edge features are extracted from each LiDAR scan by evaluating the smoothness of the local surface along the same scan line. Moreover, features in j-th scan and those in i-th map are associated using a nearest neighbor search. With this association established, the distances between the extracted features in the scan and corresponding points in the map can be calculated as follows:

$$d^e = \frac{\|(\mathbf{p}_j^e - \mathbf{p}_{i,1}^e) \times (\mathbf{p}_j^e - \mathbf{p}_{i,2}^e)\|}{\|\mathbf{p}_{i,1}^e - \mathbf{p}_{i,2}^e\|}, \tag{1}$$

$$d^p = \frac{\|(\mathbf{p}_j^p - \mathbf{p}_{i,1}^p)((\mathbf{p}_{i,1}^p - \mathbf{p}_{i,2}^p) \times (\mathbf{p}_{i,1}^p - \mathbf{p}_{i,3}^p))\|}{\|(\mathbf{p}_{i,1}^p - \mathbf{p}_{i,2}^p) \times (\mathbf{p}_{i,1}^p - \mathbf{p}_{i,3}^p)\|}, \tag{2}$$

where d^e denotes the distance between $\mathbf{p}_j^e$ and corresponding prior edge features $\mathbf{p}_{i,1}^e$ and $\mathbf{p}_{i,2}^e$. d^p denotes the distance between $\mathbf{p}_j^p$ and corresponding prior planar features $\mathbf{p}_{i,1}^p$, $\mathbf{p}_{i,2}^p$, and $\mathbf{p}_{i,3}^p$.

2) Visual Odometry: The feature point tracking and optimization-based method is adapted for our visual odometry submodule. This method effectively addresses the scale problem of monocular vision by initialization with the alignment of visual and IMU motion. After initialization, the sliding window optimization is performed for bundle adjustment, and the pose derived from the optimization is propagated with IMU measurements.

2.3 Degeneracy Detection

1) LiDAR odometry degeneracy: Most of the degenerate cases in LiDAR generally originate from the structure-less environments, such as a long corridor or a vast open field. Nevertheless, even in these scenarios, either plane or line features still exist within LiDAR's sensing range. In the case of the plane, three non-collinear points can be extracted. One translational DOF in the direction perpendicular to the plane and two rotational DOFs, except for an axis perpendicular to the plane, can be estimated by tracking the plane points. Conversely, in the case of the line, two linear points can be extracted. Two translational DOFs, except for the direction horizontal to the line, and one rotational DOF with axis horizontal to the line can be determined by tracking the line points. Therefore, we can make physical assumptions that the degeneracy would rarely occur in the 3-DOF out of the 6-DOF when a plane or line is present. Consequently, our work primarily focuses on the degeneracy of the other 3-DOF directions. Note that LiDAR SLAM is conducted with solving Levenberg-Marquardt method as follow, when $\mathbf{J}_l$ is the Jacobian matrix of $\mathbf{f}_l$, and λ is damping factor.

$$\mathbf{x} \leftarrow \mathbf{x} - \left(\mathbf{J}_l{}^\top \mathbf{J}_l + \lambda \mathrm{diag}\left(\mathbf{J}_l{}^\top \mathbf{J}_l\right)\right)^{-1} \mathbf{J}_l{}^\top \mathbf{f}_l(\mathbf{x}), \tag{3}$$

Here, eigenvalues of $\mathbf{J}_l{}^\mathbf{T}\mathbf{J}_l$ in Eq. (1) can be utilized to detect degeneracy. $[\lambda_1,\ \lambda_2,\ \lambda_3]$ are extracted as the three smallest values from the eigenvalues of $\mathbf{J}_l{}^\mathbf{T}\mathbf{J}_l$. Then, $[\lambda_1,\ \lambda_2,\ \lambda_3]$ is normalized to $[\overline{\lambda_1},\ \overline{\lambda_2},\ \overline{\lambda_3}]$. We define a non-heuristic threshold of normalized eigenvalues using the Chi-squared test. The formulation of the Chi-squared test can be written as follow, where the value of 0.103 denotes the Chi-squared value for 2-DOF at a 95% confidence level, λ_t denotes the threshold of normalized eigenvalues to solve, and the value of e_m denotes the expectation value of the minimum eigenvalue.

$$\left(\overline{\boldsymbol{\lambda}} - \mathbf{e_m}\right)^2 / \mathbf{e_m} > 0.103, \tag{4}$$

e_m is defined as 0.291 using the constraint with $\overline{\lambda_1} < \overline{\lambda_2} < \overline{\lambda_3}$. Finally, the non-heuristic threshold λ_t is decided as 0.120 from Eq. (4). If the minimum eigenvalue of $\mathbf{J}^\mathbf{T}\mathbf{J}$ is lower than λ_t, the initial guess is "switched" from the value of LiDAR odometry to visual odometry. Inversely, if the minimum eigenvalue returns to a value greater than λ_t, the system sets the initial guess as LiDAR odometry.

2) Visual odometry degeneracy: The minimum eigenvalue of the Hessian matrix of visual odometry is unstable and remains large after failure. Therefore, we adapt failure detection of visual odometry. The number of tracked features, bias changes, and positional / rotational changes between consecutive keyframes are used for failure detection. If any of these values exceed the predefined threshold, the system treats the current state as a failure. Moreover, when the failure is detected, the state of visual odometry and the system attempts re-initialization. Until successful re-initialization is achieved, the entire system relies on pure LiDAR odometry.

2.4 *Scan-to-Map Matching*

Scan-to-map matching can fail because estimations of directions to degenerate DOFs can be unstable in structure-less environments. To prevent the effect of a degenerate DOF on the optimization process, we remap Eq. (3) considering the degenerate DOF. Given $\mathbf{H}_l$ and its eigendecomposition as $\mathbf{U\Lambda U^{-1}}$, the optimization process, when the state of LiDAR odometry is well-conditioned or visual odometry fails, is as follows:

$$\delta\mathbf{x} = -\left(\mathbf{U\Lambda U^{-1}}\right)^{-1} \mathbf{J}_l{}^\top \mathbf{d}_l. \tag{5}$$

When the state of LiDAR odometry is degenerate in at least one DOF and visual odometry does not fail, the optimization process is remapped by fusing visual and LiDAR odometry in a tightly coupled way as follows:

$$\delta\mathbf{x} = \underset{\delta\mathbf{x}}{\mathrm{argmin}} \left(\left\| \underbrace{\delta\mathbf{x} + (\mathbf{H_v}^{-1}\mathbf{J_v}^\top \mathbf{d_v})}_{\mathbf{e_v}(\delta\mathbf{x})} \right\|^2 + \left\| \underbrace{\delta\mathbf{x} + \left(\mathbf{U\Lambda_p U^{-1}}\right)^{-1} \mathbf{J}_l{}^\top \mathbf{d}_l}_{\mathbf{e}_l(\delta\mathbf{x})} \right\|^2 \right), \tag{6}$$

where Λp denotes the matrix with eigenvalues removed corresponding to degenerate DOFs from $\mathbf{\Lambda}$.

When both LiDAR odometry degeneracy and visual

```
Algorithm 1: Switching Node With Degeneracy Detection.
Input: Prior status $\mathbf{T}_{k-1}$, $\mathbf{H}_l$ in (5), status buffer queue $Q_s$
       with size $n$, status of VO $S_{\text{vo}}$, differential state of
       LO $\delta\mathbf{T}^{l}_{k-1,k}$, and VO $\delta\mathbf{T}^{v}_{k-1,k}$,
Output: Final status $\mathbf{T}_k$
3-DOF normalized eigenvalues $\bar{\boldsymbol{\lambda}} = \text{eigen}_{d_1,d_2,d_3}(\mathbf{H}_l)$
if $S_{\text{vo}} == \text{fail} \vee \forall i, \bar{\boldsymbol{\lambda}}(i) \geq \bar{\boldsymbol{\lambda}}_{\boldsymbol{t}}(i)$ then
    //Use LO propagation as the initial guess
    $\mathbf{T}^{\text{init}}_k = \mathbf{T}_{k-1} \boxplus \delta\mathbf{T}^{l}_{k-1,k}$
else if check$(Q_s)$ == "Start/End to degenerate" then
    //Use an interpolation of VO and LO as the initial
    guess
    $\mathbf{T}^{\text{init}}_k = \mathbf{T}_{k-1} \boxplus \sqrt{3}\,\bar{\lambda}_1 \delta\mathbf{T}^{l}_{k-1,k} \boxplus (1-\sqrt{3}\,\bar{\lambda}_1)\delta\mathbf{T}^{v}_{k-1,k}$
else
    //Use VO propagation as the initial guess
    $\mathbf{T}^{\text{init}}_k = \mathbf{T}_{k-1} \boxplus \delta\mathbf{T}^{v}_{k-1,k}$
end if
//Scan to map matching Update $\mathbf{T}_k$ with $\mathbf{T}^{\text{init}}_k$
following (6)
Update status buffer queue Dequeue $Q_s$ and Enqueue
current status to $Q_s$.
return $\mathbf{T}_k$
```

odometry failure occur, the optimization process is executed only along the well-conditioned DOFs. In this case, the IMU preintegration significantly impacts the undetermined directions. Note that although Eq. (6) relies on MAP fusion, our switching structure ensures robustness of the multimodal system, preventing failure or degeneration of one element from affecting the overall fusion process. The entire processes in the switching node are described in Algorithm 1.

3 Experiments

In this section, the evaluation of the accuracy and robustness of the proposed method with various datasets containing sensor degeneracy is presented. Furthermore, the effectiveness of the proposed degeneracy detection is discussed.

3.1 Datasets

We prepared various datasets with various environments. Firstly, we evaluate our method in simulated datasets: Plane, Fast rotate, and Farm datasets. These datasets contain either LiDAR or visual SLAM degeneracy. Secondly, we evaluate our method in the real world and open-sourced datasets: Handheld, Multi Floor, Long Corridor, and CERBERUS DARPA subterranean challenge datasets [3]. The Handheld dataset contains degeneration of LiDAR SLAM caused by structure-less and vast open fields. As the CERBERUS dataset lacks the degeneration of LiDAR, we limit the horizontal field-of-view of LiDAR at 180° to create a more structure-less situations for each scan. The proposed method, Switch-SLAM is compared with the state-of-the-art of LiDAR, visual, and LiDAR-visual Odometry [3].

3.2 Accuracy Evaluations

As shown in Table I and Fig. 2, in most of the dataset, the proposed method shows the best performance among the compared method.

On the Fast Rotate dataset, LIO-SAM shows the best performance among the compared methods, whereas VINS-MONO fails in their localization because of

Table I : Comparison of Absolute Translational Errors on Prepared Datasets.

Dataset	Fast Rotate		Plane		Farm		Handheld		Multi Floor		Long Corridor		ANYmal 1		ANYmal 2		ANYmal 3		ANYmal 4	
	Max	RMSE	Max	RMSE	Max	RMSE	Max	RMSE	Max	RMSE	Max	RMSE	Max	RMSE	Max	RMSE	Max	RMSE	Max	RMSE
LOAM	1.41	0.44	-	-	-	-	-	-	17.9	10.6	25.6	12.6	10.26	6.63	9.67	5.05	7.81	2.39	5.79	3.90
LIO-SAM	**0.72**	**0.21**	-	-	-	-	-	-	-	-	17.6	7.64	8.38	3.77	-	-	7.10	3.52	2.47	**1.02**
VINS-MONO	-	-	**1.17**	0.41	-	-	21.4	10.3	12.8	6.30	23.8	11.8	24.9	9.08	36.9	15.1	-	-	8.55	3.52
LVI-SAM	8.82	1.82	1.82	0.69	28.8	5.75	3.27	**1.23**	-	-	8.62	4.37	5.83	2.41	9.53	3.28	-	-	6.42	3.75
R2LIVE	1.67	0.64	19.5	8.53	8.52	4.21	-	-	35.2	18.5	-	-	-	-	14.5	7.29	8.60	3.73	3.90	1.18
R3LIVE	10.1	6.43	9.01	5.84	58.6	34.7	-	-	32.4	19.0	14.5	7.63	-	-	-	-	6.77	2.06	27.1	14.0
FAST-LIVO	11.3	7.12	-	-	51.2	26.5	-	-	-	-	-	-	-	-	4.87	1.48	-	-	-	-
Switch-SLAM	1.50	0.23	1.27	**0.35**	**1.10**	**0.38**	**3.07**	1.25	**3.63**	**1.61**	**5.09**	**2.42**	**2.96**	**1.29**	**3.41**	**1.37**	**3.68**	**1.61**	**2.42**	1.05

"-" denotes the failure of localization. The units are in meters.

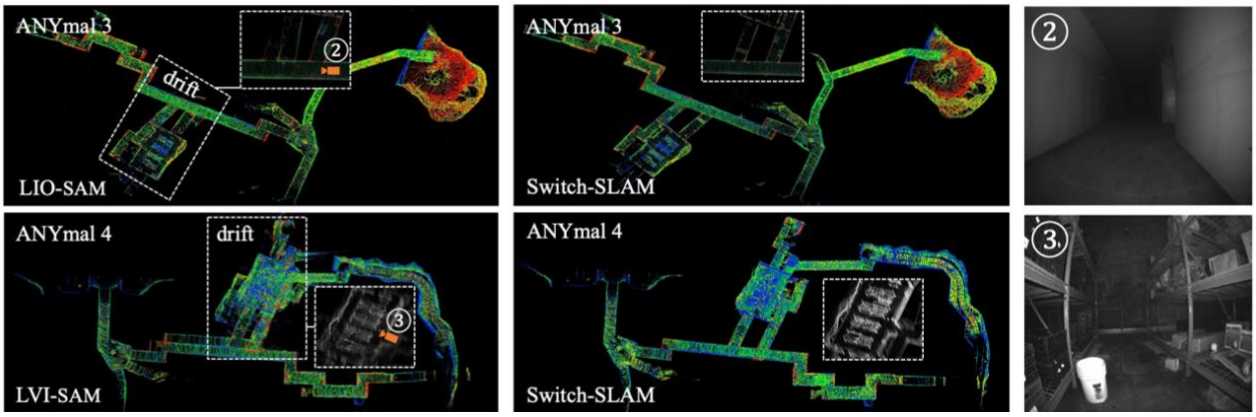

Fig. 2 Resulting maps from the compared methods and Switch-SLAM

aggressive rotation. Compared LiDAR visual inertial odometry (LVIO) methods demonstrate a larger drift than pure LiDAR-based methods. Our method is competitive with LIO-SAM because Switch-SLAM works as pure LiDAR SLAM in well-structured environments using the switching structure. On the Plane dataset, which mainly contains less-structured ground-only environments, the proposed method and VIN-MONO exhibit the best performance among the compared methods, whereas the LiDAR-based methods fail in their localization. Our method also outperforms state-of-the-art of LiDAR-visual SLAM because Switch-SLAM mainly employs visual odometry for its initial guess of scan matching in less-structured environments.

On the Farm dataset, which contains both aggressive motion and less-structured environments, LiDAR odometry fails in the phase of mapping less-structured environments, whereas visual odometry fails in the phase of aggressive motion. Conversely, the proposed method outperforms not only compared LiDAR and visual SLAM but also the state-of-the-art LVIO methods. This result is attributed to the switching structure, which allows for appropriate status transitions based on the given environmental conditions.

In the Handheld dataset, the proposed method is competitive with LVI-SAM, whereas it outperforms the other compared methods, When visual SLAM degeneracy is prolonged such as in the Fast Rotate and Farm datasets, LVI-SAM can drift significantly compared to the proposed method. On the Multi Floor and Long Corridor dataset, the proposed method shows the best performance among the compared methods. Most of the compared methods suffer with scenes featuring both structure-less environments and visual degradation. By comparison, the proposed method deals with these challenges well using the switching-based optimization as expressed in Eq. (6).

On the CERBERUS dataset, the proposed method demonstrates the best performance in ANYmal 1 and ANYmal 2. This result highlights the ability of Switch-SLAM to effectively address LiDAR degeneration, even outperforming the compared LVIO methods. In ANYmal 3, which experiences a single camera interruption, VINS-MONO and LVI-SAM fail in mapping. Moreover, the corridor-like structure makes LOAM and LIO-SAM degenerate. Conversely, Switch-SLAM successfully conducts SLAM in these environments, owing to its switching structure.

3.3 Degeneracy Detection Evaluation

To evaluate the accuracy of degeneracy detection, we compare the proposed method with the state-of-the-arts methods. The ground truth is prepared by comparing GNSS data with scan-to-scan matching using ICP at each keyframe.

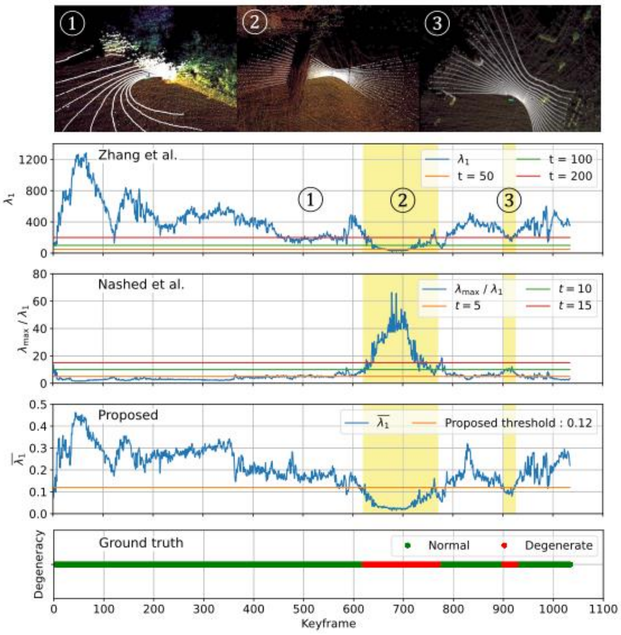

Fig. 3 Comparison of degeneracy detection performance.

The comparison of the proposed method with the state-of-the-arts is illustrated in Fig. 3. Notably, during the third phase of degeneracy, the proposed method successfully detects the degeneracy, which the state-of-the-art methods fail to identify. Note that compared methods are sensitive to threshold tuning, which is not required by our method. This detection is accomplished by normalizing the minimum eigenvalue using 3-DOF eigenvalues and applying a predefined threshold based on the Chi-squared test.

4 Conclusion

In this paper, we propose Switch-SLAM. By tackling the limitations of MAP-based sensor fusion, Switch-SLAM introduces a novel switching-based sensor fusion approach utilizing a switching structure to enhance accuracy in degenerate situations. Switch- SLAM demonstrates superior performance when compared to the state-of-the-art SLAM in terms of accuracy and localizability.

Reference

1) J. Lee, et al. : "Switch-SLAM: Switching-Based LiDAR-Inertial-Visual SLAM for Degenerate Environments, " IEEE Robotics and Automation Letters, vol 9, no. 8, pp. 7270-7277, (2024).

2) M. Tranzatto, et al.: “Cerberus in the darpa subterranean challenge,” Science Robotics, vol. 7, no. 66, p. eabp9742, (2022).

3) T. Shan, et al. : "Lvi-sam: Tightly-coupled lidar-visual-inertial odometry via smoothing and mapping," in Proceedings of the 2021 IEEE International Conference on Robotics and Automation, pp. 5692-5698, (2021).