A Loosely Coupled Vision-LiDAR Odometry using Covariance Intersection Filtering

Songming Chen, Vincent Frémont

A Loosely Coupled Vision-LiDAR Odometry using Covariance Intersection Filtering

Songming Chen[1], Vincent Frémont[1]

***Abstract*— This paper presents a loosely-coupled sensor fusion approach, which efficiently combines complementary visual and range sensor information to estimate the vehicle ego-motion. Descriptor-based and distance-based matching strategies are respectively applied to visual and range measurements for feature tracking. Nonlinear optimization optimally estimates the relative pose across consecutive frames and an uncertainty analysis using forward and backward covariance propagation is made to model the estimation accuracy. Covariance intersection filter paves the way for us to loosely couple stereo vision and LiDAR odometry considering respective uncertainties. We evaluate our approach with KITTI dataset which shows its effectiveness to fierce rotational motion and temporary absence of visual features, achieving the average relative translation error of 0.84% for the challenging 01 sequence on the highway.**

I. Introduction

Accurate state estimation and a good knowledge of the surrounding environment are crucial for autonomous driving. Many autonomous vehicles use range-based LiDAR and/or vision-based stereo cameras to perform the task of ego-motion estimation. The most frequently used sensors (Camera and LiDAR) have their own strengths and weaknesses under different working conditions. Laser scanners, for example, are good at measuring distance but are quite sensitive to fog and rain. Cameras are more commonly applied to extract the visual cues of the scene, but cannot work in low illumination conditions. Since range and visual sensors appear to be complementary, their combination allows to compensate respective shortcomings. The main challenge for long-term state estimation is error accumulation, especially in environmentally degenerate scenarios. The fusion of range and visual sensors could restrict the local uncertainties and allow to confine the odometry drift.

ORB-SLAM [1] is regraded as a typical representative of vision-based SLAM. Oriented FAST and Rotated BRIEF (ORB) features are extracted and matched for real-time state estimation, and meanwhile the bag-of-words (BOW) dictionary is queried for loop closure and drift cancellation. This method can accurately localize the mobile platform and create a sparse feature map of its surroundings with limited computation resources. ORB-SLAM2 is proposed in [2] with stereo observations based back-end which solves the scale ambiguities for trajectory estimation.

Scan-matching is a fundamental process to estimate platform motion and to create a 3D map with LiDAR. A popular approach for LiDAR based localization is LOAM [3]. It conducts Iterative Closest Point (ICP) scan-matching for 3D point clouds registration, which is followed by a global scan-to-map alignment in order to reduce local errors. Feature alignment problem can be solved using the well known Levenberg–Marquardt optimizer.

Despite the success and popularity of ORB2 SLAM [2] and LOAM [3], they are in fact deterministic algorithms. They do not effectively handle the sources of uncertainty. As a result, they provide overconfident state estimation results across frames. The main contribution of this paper is to propose a loosely coupled sensor fusion approach for vehicle localization with range and visual sensors. Measurement uncertainties for visual and range sensors are properly defined for state estimation. Backward covariance propagation [4] is utilized to transform the covariance from measurement domain to estimation domain. At the same time, forward covariance propagation is leveraged to transform the uncertainty from manifold space to Euclidean space. The covariance intersection filtering enables adaptive fusion of the two sensors given their respective uncertainties.

The remainder of this paper is divided into the following sections. Section II presents the related sensor fusion approaches to improve the perception and localization performance. The loosely coupled vision-LiDAR odometry approach is proposed in Section III and tested with the KITTI odometry benchmark in Section IV. In Section V, a concise conclusion is given which is followed by the future work plan.

* This work was supported by China Scholarship Council

[1] S. Chen, V. Frémont are with the Laboratoire des sciences du numérique de Nantes (LS2N), UMR 6004, at École Centrale de Nantes, 44321 Nantes, France, songming.chen@ec-nantes.fr, vincent.fremont@ec-nantes.fr

II. Related work

A. Enhanced Visual State Estimation

The state estimation result reached by visual-SLAM algorithms can be further enhanced via integrating LiDAR measurements. In [5], depth information from LiDAR measurements was utilized for visual feature tracking after LiDAR points being projected onto image frames. At the same time, visual semantic information was used for removing outliers and increasing the weights of static landmarks. Instead of using visual feature points, in [6] a SLAM system using visual photometric information was proposed. Its performance was enhanced with the involvement of sparse LiDAR point cloud for depth acquisition. However, as pixel resolution was much greater than LiDAR point cloud one, many pixels were not assigned the depth value, thus extra interpolation was needed to make up the missing values.

B. Enhanced LiDAR State Estimation

In many cases, LiDAR scan-matching is used for local motion estimation and visual hint is utilized for loop closure validation. The accuracy of LiDAR based localization was improved in [7], with visual feature aided loop detection to reduce the accumulated drift. In [8], the visual keyframes were utilized to assist the laser-based slam to perform local and global bundle adjustments. Furthermore, the LiDAR scan-to-scan matching can be improved using the initial guess from visual estimation as demonstrated in [9].

C. Concurrent Visual-LiDAR State Estimation

There are also many works which coupled both LiDAR and visual state estimation process together. Zhang et al. [10] designed V-LOAM pipeline which used high frequency vision based odometry as the motion prior and corrected with high precision, low frequency lidar scan matching estimation afterwards. The framework in [11] did not rely on visual estimation as the motion initial guess for lidar odometry. They took in both visual and LiDAR measurements, stacking and minimizing both modalities' residuals during the optimization phase. However, as mentioned in Section I, they did not consider the uncertainty during state estimation process, which may cause the overconfident estimation prone to certain sensor modality.

III. Presentation of the method

In Fig.1, it can be seen that the proposed sensor fusion framework starts with a descriptor-based visual feature tracking module to estimate vehicle ego-motion.

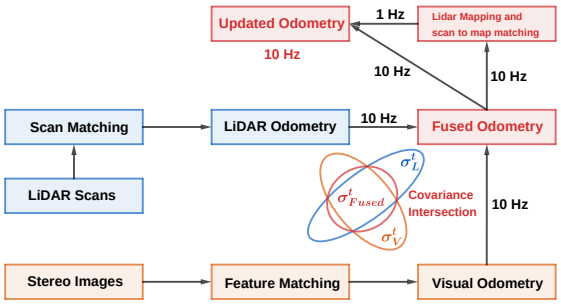


Fig. 1. Overview of the proposed sensor fusion scheme

Meanwhile, LiDAR distance-based scan-to-scan matching runs in parallel for state estimation. Backward covariance propagation transforms the uncertainty from measurement space to estimation space, which helps to obtain the uncertainty of frame-to-frame state estimation for both sensor modalities. Covariance intersection filtering ensures that the uncertainty of state does not expand after the sensor fusion, which combines two frame-to-frame poses elegantly. Then, the robustified pose is used for LiDAR point cloud registration. Scan-to-map matching afterwards further reduces local drift caused by frame-to-frame estimation. The updated odometry is the final output which is published at 10Hz frequency.

To denote coordinate systems, this paper's convention is to use uppercase letter to indicate different coordinate frames. In visualization and sensor fusion steps, vehicle pose is expressed with 3D translation and RPY Euler angle rotation. However, in order to avoid singularity problem, we optimize on their manifold which is detailed in Section III-A.1. In the following, coordinate systems being used are explained.

- Camera sensor coordinate system $\{C_t\}$ at timestamp t is defined at the camera optical center. The x-axis, y-axis and z-axis point rightward, downward and forward respectively as the camera configuration in [12].
- LiDAR sensor coordinate system $\{L_t\}$ at timestamp t is defined at the LiDAR scanner center. The x-axis, y-axis and z-axis point forward, leftward, and upward respectively as the LiDAR configuration in [12].
- World coordinate system $\{W\}$ is defined as $\{C_0\}$ which is the initial frame of the camera coordinate system, and lidar-camera extrinsics ${}^{C}T_{L}$ is assumed to be known beforehand.

A. State estimation

1) Visual odometry: In feature-based stereo vision odometry, key points with local descriptors are matched

to deduce the camera motion with scale metrics. Provided camera intrinsics $\mathbf{K}$, stereo feature points belonging to the previous frame are triangulated in the first step. And then transformed triangulated points are re-projected via the perspective projection operation $\Pr^l(\cdot)$, $\Pr^r(\cdot)$ onto the left and right images respectively considering the 6 dof state estimation variable ${}^{C_{t-1}}\hat{\Theta}_{C_t}$:

$$ {}^{C_t}\hat{\mathbf{x}}_i = \begin{bmatrix} \Pr^l\left(\mathbf{K}, {}^{C_{t-1}}\hat{\Theta}_{C_t}, {}^{C_{t-1}}\mathbf{x}_i\right) \\ \Pr^r\left(\mathbf{K}, {}^{C_{t-1}}\hat{\Theta}_{C_t}, {}^{C_{t-1}}\mathbf{x}_i\right) \end{bmatrix} \tag{1} $$

where ${}^{C_t}\hat{\mathbf{x}}_i = \left(\hat{u}_{i,l}, \hat{v}_{i,l}, \hat{u}_{i,r}, \hat{v}_{i,r}\right)^T$ is the prediction in the current frame and ${}^{C_{t-1}}\mathbf{x}_i = \left(u_{i,l}, v_{i,l}, u_{i,r}, v_{i,r}\right)^T$ is its correspondence in the previous frame. In general, the optimal relative camera transformation can be estimated by minimizing the weighted squared error of measurements and predictions.

$$ {}^{C_{t-1}}\Theta^*_{C_t} = \underset{{}^{C_{t-1}}\Theta_{C_t}}{\operatorname{argmin}} \mathbf{F}(\mathbf{x}, {}^{C_{t-1}}\Theta_{C_t}) = \underset{{}^{C_{t-1}}\Theta_{C_t}}{\operatorname{argmin}} \sum_{i=1}^{N_c} \left\| {}^{C_t}\mathbf{x}_i - {}^{C_t}\hat{\mathbf{x}}_i \right\|^2_{\Sigma} \tag{2} $$

where $\|\cdot\|^2_{\Sigma}$ is the Mahalanobis distance with $\Sigma^{-1}_{\mathbf{x}_i}$ as the information matrix for the i_{th} measurement. To handle estimation parameters that do not belong to Euclidean spaces, the common strategy is to transfer the error minimization to its corresponding manifold. In our case, iterative optimization update for estimated parameters is made using Lie algebraic perturbation model. Operator $\boxplus$ is a generalization of the normal addition operator, which is defined as $\delta\varepsilon \boxplus \hat{\Theta} \triangleq exp(\delta\varepsilon)\hat{\Theta}$, then $\tilde{\mathbf{J}}_i(\hat{\Theta})$ can be written as

$$ \tilde{\mathbf{J}}_i(\hat{\Theta}) = \left. \frac{\partial \mathbf{e}_i(\delta\varepsilon \boxplus \hat{\Theta})}{\partial \delta\varepsilon} \right|_{\delta\varepsilon \to 0} \tag{3} $$

As a result, we can apply the famous Levenberg-Marquardt method without considering additional constraint such as rotation matrix orthogonality.

2) LiDAR odometry: The same way in LiDAR odometry, edge and planar LiDAR points are tracked to recover the LiDAR pose. For each LiDAR scan point, local curvature c is computed to evaluate its smoothness considering the surrounding area. Let $\mathbb{S}$ be a group of points in the vicinity of $\mathbf{x}_i$ in the same scan layer.

$$ c = \frac{1}{|\mathbb{S}| \cdot \|{}^{L_t}\mathbf{x}_i\|} \left\| \sum_{j \in \mathbb{S}, j \neq i} \left({}^{L_t}\mathbf{x}_i - {}^{L_t}\mathbf{x}_j \right) \right\| \tag{4} $$

Edge and planar points are defined based on c values.

The edge line constructed by two edge points at previous frame $\left({}^{L_{t-1}}\mathbf{x}_j, {}^{L_{t-1}}\mathbf{x}_l\right) \in {}^{L_{t-1}}\mathbb{E}$ forms the correspondence of ${}^{L_t}\mathbf{x}_i$. ${}^{L_{t-1}}\mathbf{x}_j$ and ${}^{L_{t-1}}\mathbf{x}_j$ are selected according to nearest neighbour criteria and they belong to different scan layers to increase the point-to-line fitting robustness.

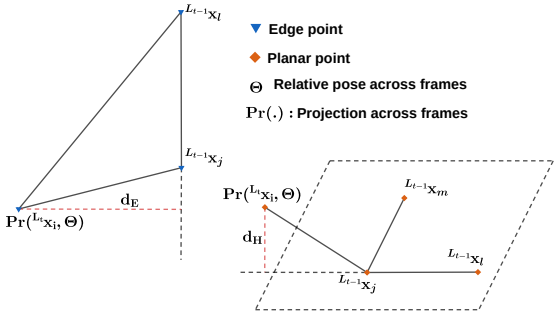


Fig. 2. Scheme of edge and planar LiDAR points correspondence projection

The planar patch represented by three points at previous frame $\left({}^{L_{t-1}}\mathbf{x}_j, {}^{L_{t-1}}\mathbf{x}_l, {}^{L_{t-1}}\mathbf{x}_m\right) \in \mathbb{H}_{t-1}$ forms the correspondence of ${}^{L_t}\mathbf{x}_i$. We assume that the closest neighbor of ${}^{L_t}\mathbf{x}_i$ is denoted as ${}^{L_{t-1}}\mathbf{x}_j$. ${}^{L_{t-1}}\mathbf{x}_l$, ${}^{L_{t-1}}\mathbf{x}_m$ are second and third nearest neighbors of ${}^{L_t}\mathbf{x}_i$, one belonging to the same scan layer of ${}^{L_{t-1}}\mathbf{x}_j$, and the other in the consecutive scan layer of ${}^{L_{t-1}}\mathbf{x}_j$. With the corresponding relationship of the feature points in hand, according to Fig. 2, we are able to calculate the distance from a feature point to its correspondence.

$$ d_{\mathbf{E}} = \frac{\left| \left({}^{L_t}\mathbf{x}_i - {}^{L_{t-1}}\mathbf{x}_j \right) \times \left({}^{L_t}\mathbf{x}_i - {}^{L_{t-1}}\mathbf{x}_l \right) \right|}{\left| {}^{L_{t-1}}\mathbf{x}_j - {}^{L_{t-1}}\mathbf{x}_l \right|} \tag{5} $$

$$ d_{\mathbf{H}} = \frac{\left| \begin{matrix} \left({}^{L_t}\mathbf{x}_i - {}^{L_{t-1}}\mathbf{x}_j \right) \\ \left(\left({}^{L_{t-1}}\mathbf{x}_j - {}^{L_{t-1}}\mathbf{x}_l \right) \times \left({}^{L_{t-1}}\mathbf{x}_j - {}^{L_{t-1}}\mathbf{x}_m \right) \right) \end{matrix} \right|}{\left| \left({}^{L_{t-1}}\mathbf{x}_j - {}^{L_{t-1}}\mathbf{x}_l \right) \times \left({}^{L_{t-1}}\mathbf{x}_j - {}^{L_{t-1}}\mathbf{x}_m \right) \right|} \tag{6} $$

The optimal LiDAR relative pose can be obtained by minimizing the weighted sum squared distances of edge and planar points to their correspondences. $\Sigma_{\mathbf{x}^{-1}_{Ei}}$ and $\Sigma_{\mathbf{x}^{-1}_{Hi}}$ stand for the information matrix of $E_{i_{th}}$ edge and $H_{i_{th}}$ planar measurement points and we take the same optimization strategy as in Section III-A.1.

$$ {}^{L_{t-1}}\Theta^*_{L_t} = \underset{{}^{L_{t-1}}\Theta_{L_t}}{\operatorname{argmin}} \sum_{E_i=1}^{N_E} d_{E_i} \Sigma_{\mathbf{x}^{-1}_{Ei}} d_{E_i} + \sum_{H_i=1}^{N_H} d_{H_i} \Sigma_{\mathbf{x}^{-1}_{Hi}} d_{H_i} \tag{7} $$

B. Uncertainty analysis

Robust state estimation should be able to provide the uncertainty information associated with the vehicle pose estimates. The sensor fusion phase is driven by the uncertainties in the estimation domain. Thus, we analyse the uncertainties coming from visual and range sensors via forward and backward covariance propagation.

1) Visual sensor uncertainty: Although the optimal relative pose can be obtained by minimizing Eq. 2, its accuracy also depends on the precision of the corresponding feature points, more specifically, the level of the image pyramid they belong to. The image pyramid [13] is a series of image collections whose resolution gradually decreases in the shape of a pyramid. The image pyramid can be sequentially matched to ensure scale invariance during feature tracking. In our case, the image pyramid has 8 levels with the same scale factor 1.2 between two consecutive levels. We assume that all points considered in the optimization procedure are well-matched pixel features with only zero mean Gaussian noise $N(0,\Sigma^V_{\mathbf{x}_i})$, with $\sigma_{\mathbf{x}_{u_{i,l(r)}}} = \sigma_{\mathbf{x}_{v_{i,l(r)}}} = 1.2^{level-1}$ as standard deviation for i_{th} measurement. Jacobian matrix $\tilde{\mathbf{J}}_i(\Theta^*)$ is defined in Eq. 3, and it converts the uncertainty from measurement space to estimation space. Since we optimize on manifold, let $\varepsilon = log(\Theta^*)$ a 6d vector in the Lie algebra space, Jacobian matrix $\mathbf{J}_{m2e} = \frac{\partial e^{\varepsilon}}{\partial \varepsilon}$ is indispensable to propagate the covariance from manifold space to Euclidean space for data visualization and fusion. Then we can obtain the uncertainty of frame-to-frame relative pose $\Sigma^V_{\Theta^*}$ through Eq. 8 and the result is shown in Fig. 3.

$$\Sigma^V_{\Theta^*} = \mathbf{J}^V_{m2e}\Big(\sum_{i=1}^{N_c}(\tilde{\mathbf{J}}_i^{V\prime}(\Theta^*)\Sigma^{V\,-1}_{\mathbf{x}_i}\tilde{\mathbf{J}}^V_i(\Theta^*))\Big)^{-1}\mathbf{J}^{V\;\prime}_{m2e} \quad (8)$$

2) Range sensor uncertainty: In our case, a Velodyne HDL-64E is used which provides a $(0^\circ \sim 360^\circ)$ azimuth field of view (θ) and $(-24.9^\circ \sim 2^\circ)$ elevation field of view (ϕ). According to official velodyne data sheet, range accuracy can reach up to 2 cm which is quite small compared with its range limit 120 m. Hence, each measurement is treated equally and we can set measurement uncertainty $\Sigma^L_{\mathbf{x}_i}$ as identity matrix for each point. Based on such assumption, we obtain the uncertainty of scan-to-scan relative pose $\Sigma^L_{\Theta^*}$ through Eq. 9 and the result is shown in Fig. 4.

$$\Sigma^L_{\Theta^*} = \mathbf{J}^L_{m2e}\Big(\sum_{i=1}^{N_E+N_H}(\tilde{\mathbf{J}}_i^{L\prime}(\Theta^*)\Sigma^{L\,-1}_{\mathbf{x}_i}\tilde{\mathbf{J}}^L_i(\Theta^*))\Big)^{-1}\mathbf{J}^{L\;\prime}_{m2e} \quad (9)$$

3) Covariance Intersection Filtering: Covariance intersection [14] is a variant of Gaussian process sensor fusion which can combine two estimates under unknown correlations. The covariance intersection combination formulas are given by

$$\begin{aligned}\Sigma^{fused}_{\Theta} &= \left(\omega\left(\Sigma^L_{\Theta}\right)^{-1} + (1-\omega)\left(\Sigma^V_{\Theta}\right)^{-1}\right)^{-1}\\ \Theta^{fused} &= \Sigma^{fused}_{\Theta}\left(\omega\left(\Sigma^L_{\Theta}\right)^{-1}\Theta^L + (1-\omega)\left(\Sigma^V_{\Theta}\right)^{-1}\Theta^V\right)\end{aligned} \quad (10)$$

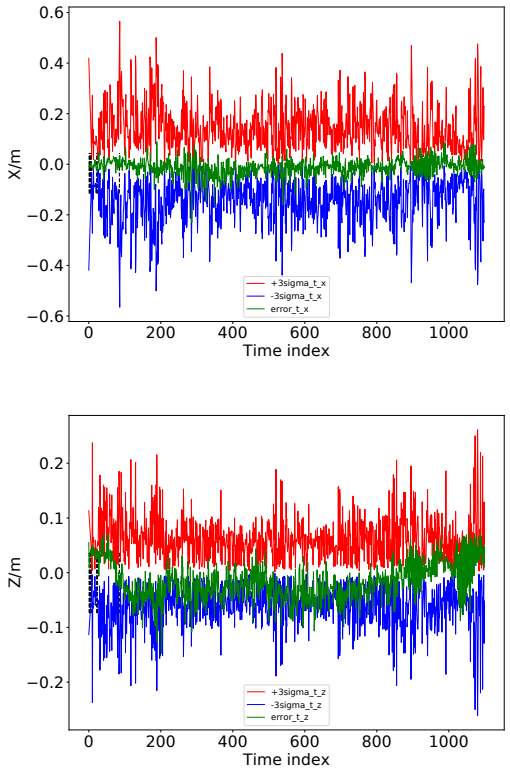


Fig. 3. Visual sensor pose estimation uncertainty along t_x and t_z for KITTI sequence 01

where $\omega \in [0,1]$ minimizes trace of the fused covariance matrix Σ^{fused}_{Θ} at each step. If the Jacobian matrix is near singular, probably because of local minimal occurrence or individual sensor failure, then inverting $\tilde{\mathbf{J}}_i'(\Theta^*)\Sigma^{-1}_{\mathbf{x}_i}\tilde{\mathbf{J}}_i(\Theta^*)$ will lead to unreliable uncertainty estimation marked as black dash lines in Fig. 3. Covariance intersection can ensure that the resulting estimate is conservative, which efficiently filters out the unstable estimation. In order to simplify the fusion parameterization, only planar translation t_x and t_z and yaw angle r_y are fused and updated. The fused pose will better register the lidar map point and we adopt the multi-level voxel scan-to-map matching as in [11] to reduce the frame-to-frame estimation drift.

IV. EXPERIMENTAL RESULTS

The KITTI dataset [12] contains stereo sequences and Velodyne HDL-64E LiDAR point clouds captured in urban and highway environments. We use the metric of average relative translation error t_{rel} proposed in [12] for evaluation purpose. To have a fair comparison, the ORB SLAM2 loop closure module is deactivated.

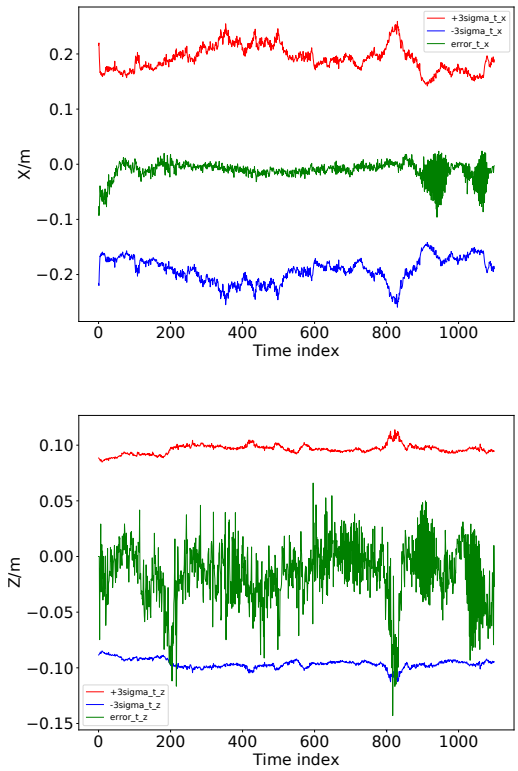


Fig. 4. Range sensor pose estimation uncertainty along t_x and t_z for KITTI sequence 01

Evaluation results are obtained on a laptop with an Intel i7-9750H CPU and 32GB of RAM. We choose three typical sequences 01 (Highway), 02 (Urban+Country) and 07 (Urban) from KITTI dataset to make analysis and detailed quantitative result is shown in Tab. I [1].

Our Loosely-Coupled Vision-LiDAR Odometry(LC-VLO) outperforms state-of-the art approaches for challenging trajectory in sequence 01. When driving on a highway scenario, few distinctive visual features are available, see Fig. 5, which makes descriptor-based feature tracking erroneous and thus causes poor pose estimation for visual sensor. Our proposed approach ensures a consistent odometry estimation even moving at high speed. In sequence 02, our loosely coupled odometry is not as good as the ORB-SLAM2 due to lack of horizontal lines or planes to constraint the drift along the vertical axis. However, it does efficiently prevent large divergence occurrence like in A-LOAM[2] method. The large divergence mainly results from A-LOAM's inappropriate distance-based matching strategy. Far edge points are more likely to be mismatched when encountering large rotational motion. It happens at the middle of sequence 02 where A-LOAM method loses the tracking of features and fails to confine the estimation error. Uncertainty analysis is able to detect the potential deficiencies in the early scan-to-scan step and mitigate feature misalignment problem. Our LC-VLO is superior to ORB-SLAM2 and A-LOAM for sequence 7, which shows that multi-level voxel scan-to-map matching procedure is indispensable to reduce frame-to-frame estimation drift. Overall, the proposed LC-VLO adaptively fuses vision and LiDAR estimation, which is able to improve estimation performance in individual sensor degenerate cases, especially for the challenging KITTI sequence 01 and 02.

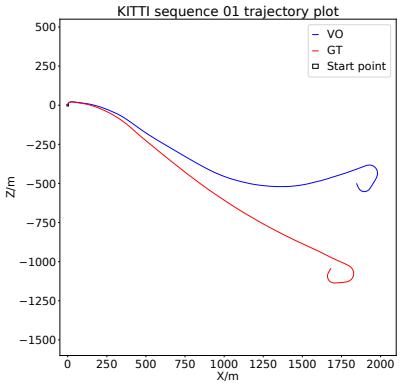


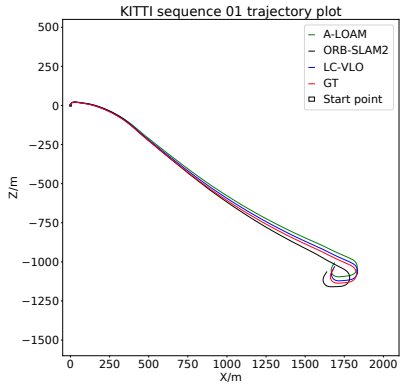
Fig. 5. Few distinctive ORB visual features for tracking on the highway scenario, KITTI sequence 01

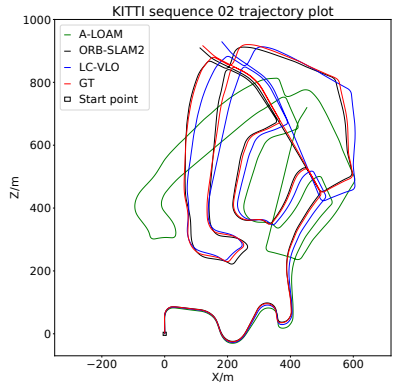
V. Conclusion and Future Work

In this paper, we use covariance intersection filtering to robustify the odometry estimation from two data streams. The effectiveness of the proposed method has been verified on the public KITTI VO benchmark. The result shows its robustness to large rotational motion and temporary absence of visual features as a result of our anisotropic uncertainty modelling in the sensor fusion step. Since we perform the sensor fusion in a loosely-coupled manner, each sensor modality can

[1]The sequence 08 is not evaluated due to ground truth flaw with manual inspection

[2]Advanced implementation of LOAM, https://github.com/HKUST-Aerial-Robotics/A-LOAM





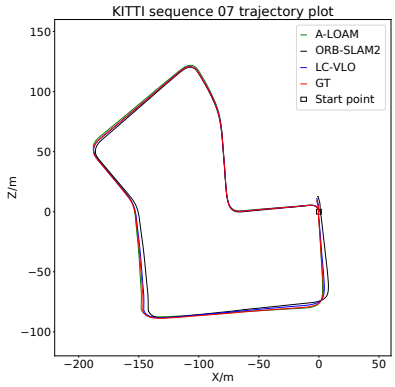


Fig. 6. Estimated trajectory and ground-truth for KITTI 01, 02 and 07 sequences

be easily replaced according to personalized demands, which makes our approach very flexible. As our current approach does not consider loop closure, we will focus on exploiting visual semantic hints for robust feature tracking and place recognition to further ameliorate the localization accuracy in the future.

TABLE I
COMPARISON OF ACCURACY (IN PERCENTAGE).

Sequence \ Metric t_{rel}	ORB-SLAM2	A-LOAM	LC-VLO
00	0.88%	0.77%	**0.74**%
01	1.40%	2.27%	**0.84**%
02	**0.79**%	4.91%	1.50%
03	**0.77**%	1.24%	0.87%
04	**0.45**%	1.23%	1.08%
05	0.61%	0.70%	**0.43**%
06	0.73%	0.62%	**0.58**%
07	0.90%	0.63%	**0.50**%
08	–%	–%	–%
09	**0.90**%	1.09%	1.01%
10	0.59%	1.69%	**0.52**%

REFERENCES

[1] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[2] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[3] J. Zhang and S. Singh, "Loam: Lidar odometry and mapping in real-time." in *Robotics: Science and Systems*, vol. 2, no. 9, 2014.

[4] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[5] J. Graeter, A. Wilczynski, and M. Lauer, "Limo: Lidar-monocular visual odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 7872–7879.

[6] Y.-S. Shin, Y. S. Park, and A. Kim, "Direct visual slam using sparse depth for camera-lidar system," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–8.

[7] X. Liang, H. Chen, Y. Li, and Y. Liu, "Visual laser-slam in large-scale indoor environments," in *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2016, pp. 19–24.

[8] Z. Zhu, S. Yang, H. Dai, and F. Li, "Loop detection and correction of 3d laser-based slam with visual information," in *Proceedings of the 31st International Conference on Computer Animation and Social Agents*, 2018, pp. 53–58.

[9] G. Pandey, S. Savarese, J. R. McBride, and R. M. Eustice, "Visually bootstrapped generalized icp," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 2660–2667.

[10] J. Zhang and S. Singh, "Visual-lidar odometry and mapping: Low-drift, robust, and fast," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 2174–2181.

[11] Y. Seo and C.-C. Chou, "A tight coupling of vision-lidar measurements for an effective odometry," in *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 1118–1123.

[12] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[13] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," *RCA engineer*, vol. 29, no. 6, pp. 33–41, 1984.

[14] S. J. Julier and J. K. Uhlmann, "Using covariance intersection for slam," *Robotics and Autonomous Systems*, vol. 55, no. 1, pp. 3–20, 2007.