ROBUST OBJECT-AWARE SAMPLE CONSENSUS WITH APPLICATION TO LIDAR ODOMETRY

Hui Cheng, Yongheng Hu, Chongyu Chen, and Liang Lin

School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, Guangdong, 510006 China.

ABSTRACT

Random sample consensus (RANSAC) is a popular paradigm for parameter estimation with outlier detection, which plays an essential role in 3D robot vision, especially for LiDAR odometry. The success of RANSAC strongly depends on the probability of selecting a subset of pure inliers, which sets barriers to robust and fast parameter estimation. Although significant efforts have been made to improve RANSAC in various scenarios, its strong dependency on inlier selection is still a problem. In this paper, we propose to address such dependency in the context of LiDAR odometry by robust object-aware sample consensus (ROSAC). In the proposed ROSAC, the sampling strategy is adjusted to preserve object shapes and a new consensus method is developed based on robust low-dimensional subspace analysis. It is demonstrated in extensive experiments that the proposed paradigm works well in LiDAR odometry, achieving estimation of 3D pose with superior accuracy compared to RANSAC. Even for the case of RANSAC failure, ROSAC still achieves up to 67% of improvement in accuracy compared to baseline LiDAR odometry. Since a partially parallel implementation of ROSAC already leads to a significant speedup, we believe it can be extended to other problems of parameter estimation with both higher accuracy and efficiency.

***Index Terms*—** RANSAC, Low-rank and sparse decomposition, Robust estimation, LiDAR odometry

1. INTRODUCTION

Random sample consensus (RANSAC) [1] is a famous paradigm for parameter estimation with outlier rejection, which has been successfully applied to solve many computer vision and signal processing problems. In 3D computer vision, RANSAC is usually adopted with iterative closest point (ICP) methods [2, 3, 4] to estimate the camera pose while rejecting outliers. But it produces correct results only with a certain probability because it depends on the probability of selecting a subset of pure inliers. Although variants of RANSAC have been proposed for various applications [5], but their improvements upon RANSAC in robustness are very limited. Since the outlier distribution changes with the sensor type, the probability of selecting pure inliers is also dependent on the utilized sensor. It is non-trivial to apply RANSAC to deal with real data [6], especially for the cases without strict correspondences such as LiDAR odometry that estimate the 3D pose from range measurements.

It is well-known that LiDAR range measurements are accurate but sparse, especially along the vertical direction (e.g. 16 lines of data per frame). Such spatial sparsity leads to measurement drifts on object surfaces (e.g., a scanning line may be on a person's head in the first frame and then the person's neck in the next frame). Therefore, the success rate of classical RANSAC becomes very low and researchers tend to adopt extra components for handling outliers in LiDAR odometry. Representative techniques include object detection [7, 8] and tracking [9, 10, 11]. Although being more robust against moving objects, these methods inevitably desire more computations due to the additional components. More importantly, the strong dependency of selecting pure inliers in RANSAC remains unsolved.

Robust subspace analysis is an important tool to reveal the intrinsic characteristics from a data matrix. Representative algorithms include the truncated robust principal component analysis (RPCA) [12, 13, 14] that recovers the low-rank subspace of the data corrupted by non-Gaussian noise and random missing. These algorithms have been successfully applied to analysis of sequential data such as color videos [15] and RGB-D (synchronized color and depth) sequences [16]. For the improving subspace analysis by sample consensus algorithms, there have been several pioneer trials [17, 18]. Moreover, the trials of improving sample consensus by subspace analysis are quite limited.

Inspired by the recent advance in robust subspace analysis, we propose a new paradigm for sample consensus that does not rely on the selection of pure inliers. In particular, the samples are divided into several subsets with certain overlapping. A number of model parameters (i.e., rigid transforms in the case of LiDAR odometry) are estimated using these subsets individually. Then, every parameter is regarded as a point in a multi-dimensional space. Based on the observation that such multi-dimensional points essentially lie on a

This work is partially supported by the National Natural Science Foundation of China (NSFC) under Grant 61602533, NSFC-Shenzhen Robotics Projects (U1613211), The Fundamental Research Funds for the Central Universities, and Science and Technology Program of Guangzhou, China (201510010126). Correspondence should be addressed to Chongyu Chen chenchy47@mail.sysu.edu.cn.

low-dimensional manifold defined by the model parameter, we perform robust subspace analysis to extract the outliers of these points, resulting in a number of inlier parameters. The final parameter is given by averaging all inlier parameters. In this way, we address the dependency of RANSAC on selecting good subsets. Since every subset can be processed individually, the proposed paradigm is naturally with the ability of being implemented in parallel. It is believed that the proposed paradigm can lead to more efficient solutions to real-time 3D robot vision using consumer-grade devices. Taking the application of LiDAR odometry as an example, we focus on the estimation of rigid transforms in the rest of this paper.

2. PROPOSED PARADIGM

The main idea of the proposed paradigm is based on two observations. First, the rigid transforms estimated using different subsets from the same LiDAR frame can be seen as samples from the same low-dimensional manifold, especially when the LiDAR device is with rotations and translations of limited degrees of freedom. Second, the estimated rigid transforms are typically with errors, which can be formulated as Gaussian-like noise and sparse outliers that can be well handled by robust subspace analysis.

2.1. Overall Pipeline

Similar to RANSAC, the proposed paradigm consists of two main stages, i.e., sampling stage and sample consensus stage. Given two frames of range measurements P and P' with N_p points per frame, we use the following 4 steps for estimating the rigid transform with sample consensus:

1. Select N_t pairs of subsets from P and P';
2. Estimate the rigid transform between every subset pair individually, resulting in N_t rigid transforms $TF = \{TF_1, TF_2, \ldots, TF_{N_t}\}$;
3. Determine the outlier rigid transforms from TF using robust subspace analysis;
4. Estimate the final transform by averaging the inlier rigid transforms from TF.

One can see that the proposed paradigm integrates object-aware sampling and robust subspace analysis. Thereby, we name it Robust Object-aware SAmple Consensus (ROSAC). Note that there are two configurations for step 1), respectively refer to the "single-sided" and "double-sided" sampling strategies. The "single-sided" sampling is to select N_t subsets $S = \{S_1, S_2, \ldots, S_{N_t}\}$ of n $(n < N_p)$ points from P. Then, the whole frame P' and S_i is considered as one pair of subset, whose rigid transform is estimated individually. The "double-sided" sampling is to perform object-aware selection on both P and P', resulting in N_t pairs of small subsets, i.e., $\{S_1, S'_1\}, \ldots,$ and $\{S_N, S'_N\}$.

The subset selection in step 1) and the outlier extraction in step 3) will be described in Sec. 2.3 and Sec. 2.2, respectively. It should be noted that steps 1 and 2 can be implemented in parallel for faster calculation. Since N_t can is usually very small (e.g. $N_t = 20$). Such parallel implementation can be used on devices with either multi-core CPU or low-cost GPU.

2.2. Object-aware Subsampling

Subsampling is a common technique for reducing computational cost while keeping reasonable scanning fidelity, especially in the processing of 3D points [19]. Random sampling and uniform sampling are two commonly used strategies for selecting subsets, which may not be an ideal choice for LiDAR measurements. The main reason is that, different from the point cloud registration in other scenarios which is dense, accurate, and even with point-to-point correspondences, LiDAR measurements are sparse, noisy, and without any correspondence. More severely, there may be no point-to-point correspondence between two sequential LiDAR frames because the probability of a same point on the object scanned by the laser twice is very small, especially for the mainstream LiDAR devices that only have 16, 32 or 64 scanning lines.

In the proposed paradigm, we propose to combine uniform sampling with object-aware preservation for sampling the LiDAR measurements. Two issues are considered in such combination. First, uniform sampling is the best way of maintaining the global structure of LiDAR scans. Second, it is observed that thin objects only occupy limited number of horizontal measurements in LiDAR frames. If uniform sampling is performed, there will be very few measurements on such objects, which probably act as outliers and deteriorate the registration accuracy. Therefore, to avoid such deterioration, we propose to preserve tiny objects by boundary detection. In particular, we compute the difference between neighboring measurements. For two very close boundaries, (i.e., their interval is not larger than N_d measurements), the measurements between them are always kept in every subset.

2.3. Robust Sample Consensus

With the object-aware subsets S, one can perform any registration method on S and P' to estimate N rigid transforms TF. Then, we divide the i-th rigid transform TF_i into two parts, i.e., the rotation angles $\Theta_i = [\theta_x^{(i)}, \theta_y^{(i)}, \theta_z^{(i)}]$ and the translation vector $T_i = [T_x^{(i)}, T_y^{(i)}, T_z^{(i)}]$. By regarding every part as one data point in 3D space, we construct two data matrices

$$M_\Theta = [\Theta_1^\top - \Theta_0^\top, \Theta_2^\top - \Theta_0^\top, \ldots, \Theta_{N_t}^\top - \Theta_0^\top] \quad (1)$$

and

$$M_T = [T_1^\top - T_0^\top, T_2^\top - T_0^\top, \ldots, T_{N_t}^\top - T_0^\top], \quad (2)$$

where

$$\Theta_0 = \frac{1}{N_t} \sum_{i=1}^{N_t} \Theta_i$$

and

$$T_0 = \frac{1}{N_t} \sum_{i=1}^{N_t} T_i$$

are the mean of matrix columns. Therefore, these matrices are centralized with respect to the mean of the data points.

By performing low-rank and sparse decomposition (LRSD) [14] on M_Θ and M_T individually, we obtain their low-rank and sparse components. That is,

$$\{L_\Theta, S_\Theta\} = \text{LRSD}(M_\Theta, r, \tau_\Theta), \tag{3}$$

and

$$\{L_T, S_T\} = \text{LRSD}(M_T, r, \tau_T), \tag{4}$$

where L_Θ and L_T are respectively the low-rank components of M_Θ and M_T, S_Θ and S_T are respectively the low-rank components of M_Θ and M_T, r is the target rank, τ_Θ is the threshold of outlier perturbation in M_Θ, and τ_T is the threshold of outlier perturbation in M_T. According to our experience, $r = 1$ is a reasonable choice for LiDAR odometry. The parameters τ_Θ and τ_T are suggested to be the standard deviation of M_Θ and M_T, respectively. If a non-zero value appears in the i-th column of S_Θ or S_T, the i-th rigid transform TF_i is considered as the outlier. Let $\{TF_1, \ldots, TF_p\}$ be the p inliers determined after checking the non-zero entries in S_Θ and S_T. Then, the final rigid transform TF^* is given by

$$TF^* = \frac{1}{p}\sum_{i=1}^{p} TF_i. \tag{5}$$

3. RESULTS

In this section, we conduct experiments on the commonly used KITTI dataset of odometry [20]. First, a small subset of the dataset is selected to analyze the components of the propose paradigm. Then, quantitative evaluations on both accuracy and efficiency are reported to verify the effectiveness of the propose paradigm.

This paper concentrates on the estimation of rigid transforms. Therefore, we adopt the evaluation metrics the same as that used in the KITTI dataset. That is, given the ground-truth (GT) transform matrix $T \in \mathbb{R}^{4\times 4}$ and the estimated one $\hat{T}$, the translation error e_t and rotation error e_θ are computed by

$$\begin{cases} e_t = & \|\Delta\mathbf{t}\| \\ e_\theta = & \arccos(0.5(tr(\Delta R) - 1)) \end{cases} \tag{6}$$

where $tr(\Delta R)$ represents the trace of ΔR, ΔR and $\Delta\mathbf{t}$ are defined as

$$\begin{bmatrix} \Delta R & \Delta\mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} = (\hat{T})^{-1} T. \tag{7}$$

It can be seen that the translation error is essentially the root-mean-square-error (RMSE) and the rotation error is computed in the angle domain. Thus, the units of the above metrics are generally meter and radian for translation and rotation, respectively. For the analysis of the proposed sampling strategy, we extract a test subset consisting of 103 pairs of LiDAR frames in KITTI dataset. The testing results are obtained by applying the proposed paradigm with classical ICP [21] on the test subset. The LRSD based sample consensus method is used in all tests. Key parameters of the proposed paradigm are the number of sampling times N_t and the number of points of every subset N_s.

3.1. Sampling strategy

For the analysis of the sampling strategies, we empirically choose $N_t = 20$ and $N_s = 0.05N_p$ and apply random sampling, uniform sampling, and object-aware sampling on the LiDAR measurements for the estimation of rigid transform. Table 1 presents the translation error with boldface and underline indicating the best results in the same row and column respectively. It is clearly shown that object-aware sampling consistently outperforms other alternatives in the context of LiDAR odometry with sample consensus, regardless of changing the inlier ratio of ICP. The improvements of estimation accuracy of our method upon random sampling and uniform sampling are up to 33.05% and 24.13%, respectively. Since it is observed that the propose paradigm achieves nice performance when the inlier ratio is 95%. We keep such setting in the rest of our experiments.

Inlier ratio	Random	Uniform	Object-aware
100%	0.0756	0.0665	**0.0536**
95%	0.0683	0.0622	**0.0457**
90%	0.0676	0.0653	**0.0566**
85%	0.0710	0.0696	**0.0540**

Table 1. Mean translation errors of different sampling strategies for pose estimation on a subset of the KITTI dataset.

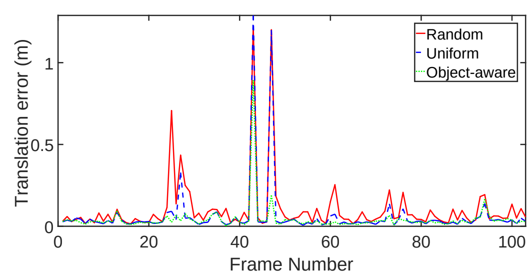

Fig. 1. Per frame translation errors on a subset of the KITTI dataset.

Per frame translation errors on the test subset are reported in Fig. 1. It can be seen that object-aware sampling performs more stable than other two sampling strategies, especially for the frames that cause registration failures of other strategies (e.g., frames #20 to #30). Better robustness is the main reason of higher accuracy of the proposed paradigm.

3.2. Key parameters

$N_t \backslash N_s$	$0.1N_p$		$0.05N_p$		$0.01N_p$		$0.005N_p$	
	e_t	e_θ	e_t	e_θ	e_t	e_θ	e_t	e_θ
10	0.0656	0.0030	0.0653	0.0031	0.0477	0.0034	0.0578	0.0037
20	-	-	0.0641	0.0031	0.0455	0.0034	0.0540	0.0036
30	-	-	-	-	0.0447	0.0034	0.0516	0.0036
50	-	-	-	-	0.0427	0.0034	0.0522	0.0036
100	-	-	-	-	0.0415	0.0034	0.0540	0.0036

Table 2. The performance changes with the key parameters.

Fixing the sampling method as object-aware sampling, we turn to evaluate its performance changes with the parameters

N_t and N_s. Larger N_t means more data points and thus better accuracy can be expected. The number of points in each subset N_s is related to N_t for the case of sampling without replacement, which directly affects the execution time of the whole algorithm. By increasing N_t from 10 to 100 and N_s from $0.005N_p$ to $0.1N_p$, we obtain the results shown in Table 2. It can be seen that our ROSAC is insensitive to the parameter changes. The accuracy difference is small when N_s is fixed. This is because there is a uniform sampling in the proposed object-aware sampling. The same set of points will appear for multiple times when N_t is greater than N_p/N_s. It is found that given a fixed N_t, the accuracy is not monotone changing with N_s.

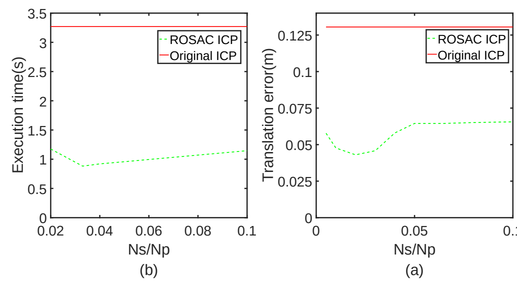

Fig. 2. The relationship between performance and parameters.

To further investigate the effect of N_s on the proposed method, we carefully assess the relationships among N_s, execution time and accuracy on a computer with a 32-core CPU (Intel Xeon E52620V4). Detailed analysis is shown in Fig. 2, in which the accuracy and efficiency curves of ICP with ROSAC are compared with that of classical point-to-plane ICP. The green curves are obtained by using a simple parallel implementation of the "double-sided" ROSAC ICP based on the MATLAB "parfor" command, and the red curves are obtained by original ICP. A speedup of up to 4× brought by the ROSAC is validated by these curves. From Fig. 2, one can see the execution time of ROSAC ICP keeps reducing as N_s decreases. Note that such accelerations stop when $N_s/N_p = 1/32$ because our CPU supports 32 threads at most, which means that we only implement a partial parallelization. It can be expected that more threads or GPU implementation will lead to higher efficiency. Meanwhile, the registration accuracy of ROSAC ICP is always higher than the original ICP, which demonstrates ROSAC can lead to both higher efficiency and accuracy in LiDAR odometry.

Sequence	ROSAC ICP		RANSAC ICP		Original ICP	
	e_t	e_θ	e_t	e_θ	e_t	e_θ
1	**0.3480**	**0.0036**	1.5431	0.0075	0.9474	0.0046
2	**0.0835**	**0.0020**	0.2590	0.0089	0.1510	0.0021
3	0.0212	0.0037	0.0593	0.0170	**0.0154**	**0.0036**
4	**0.0294**	0.0088	0.5878	0.0115	0.4345	**0.0087**
5	**0.0140**	0.0018	0.0263	0.0028	0.0146	**0.0017**
6	**0.0161**	0.0027	0.0507	0.0035	0.0219	**0.0026**
7	0.0156	0.0029	0.0234	0.0039	**0.0145**	**0.0028**
Average	**0.0754**	**0.0036**	0.3642	0.0059	0.2285	0.0037

Table 3. The translation and rotation errors on the KITTI dataset.

3.3. Evaluations on KITTI dataset

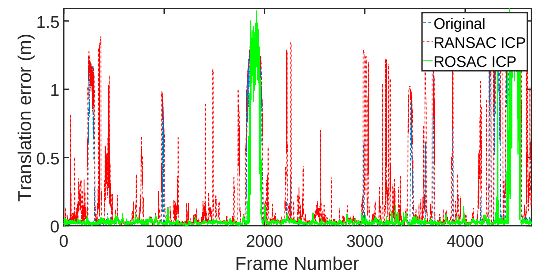

Fig. 3. Per frame translation error on sequence #2.

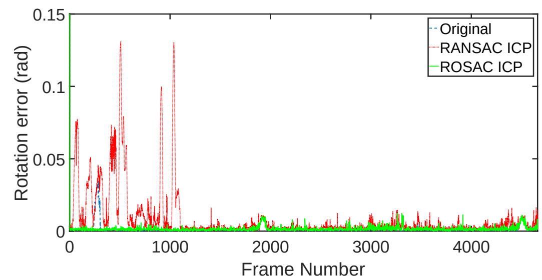

Fig. 4. Per frame rotation error on sequence #2.

For a comprehensive assessment of our method, we conduct quantitative evaluations on the whole KITTI dataset for LiDAR odometry. For a fair comparison, we use $N_t = 20$ and $N_s = 0.05N_p$ for both ROSAC and RANSAC. Estimation errors are shown in Table 3 and the visualization of these errors on a representative sequence is shown in Fig. 3 and 4. It is shown in Table 3 that RANSAC fails to produce correct estimation due to the small number of N_s, while ROSAC succeeds in greatly improving the translation accuracy and the rotation stability. To specific, ROSAC achieves improvements up to 67.00% in translation accuracy compared to the baseline ICP. The superior stability of ROSAC is validated in the per frame registration errors on sequence #2 of the dataset, as shown in Fig. 3 and Fig. 4.

4. CONCLUSIONS

In this paper, we have presented a novel paradigm for sample consensus in the context of LiDAR odometry. The proposed paradigm is based on the combination of object-aware sampling and low-rank subspace analysis techniques. The proposed paradigm can be seen as a new solution to robust parameter estimation with outlier rejection. Although the application is chosen as LiDAR odometry, the proposed paradigm is believed to be universal due to the generality of RANSAC. It would be interesting to investigate special sampling methods for more scenarios where RANSAC is applicable and find theoretical bounds for the proposed ROSAC.

5. REFERENCES

[1] Martin A Fischler and Robert C Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," , no. 6, pp. 381–395, 1981.

[2] C. Langis, M. Greenspan, and G. Godin, "The parallel iterative closest point algorithm," in *International Conference on 3-D Digital Imaging and Modeling, 2001. Proceedings*, 2001, pp. 195–202.

[3] D. Chetverikov, D. Svirko, D. Stepanov, and Pavel Krsek, "The trimmed iterative closest point algorithm," in *International Conference on Pattern Recognition, 2002. Proceedings*, 2002, pp. 545–548 vol.3.

[4] Ayush Dewan, Tim Caselitz, Gian Diego Tipaldi, and Wolfram Burgard, "Motion-based detection and tracking in 3D LiDAR scans," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 4508–4513.

[5] D. R. Myatt, Philip H. S. Torr, Slawomir J. Nasuto, J. Mark Bishop, and R. Craddock, "NAPSAC: High Noise, High Dimensional Robust Estimation - it's in the Bag," in *British Machine Vision Conference 2002, BMVC 2002, Cardiff, Uk, 2-5 September*, 2002, pp. 458–467.

[6] Charles V Stewart, "Robust parameter estimation in computer vision," *SIAM Review*, vol. 41, no. 3, pp. 513–537, 1999.

[7] T. D. Vu, J. Burlet, and O. Aycard, "Grid-based localization and online mapping with moving objects detection and tracking: new results," in *2008 IEEE Intelligent Vehicles Symposium*, June 2008, pp. 684–689.

[8] François Pomerleau, Philipp Krüsi, Francis Colas, Paul Furgale, and Roland Siegwart, "Long-term 3d map maintenance in dynamic environments," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 3712–3719.

[9] Frank Moosmann and Christoph Stiller, "Joint self-localization and tracking of generic objects in 3D range data," in *2013 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2013, pp. 1146–1152.

[10] Shao-Wen Yang and Chieh-Chih Wang, "Simultaneous egomotion estimation, segmentation, and moving object detection," *Journal of Field Robotics*, vol. 28, no. 4, pp. 565–588, 2011.

[11] Wen Xiao, Bruno Vallet, Konrad Schindler, and Nicolas Paparoditis, "Simultaneous detection and tracking of pedestrian from panoramic laser scanning data," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 295–302, 2016.

[12] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, Jun. 2011.

[13] T. Zhou and D. Tao, "GoDec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Int'l Conf. Mach. Learning (ICML)*, 2011.

[14] T. Zhou and D. Tao, "Greedy bilateral sketch, completion & smoothing," in *Int'l Conf. Artificial Intell. Statistics*, 2013.

[15] Chongyu Chen, Jianfei Cai, Weisi Lin, and Guangming Shi, "Incremental low-rank and sparse decomposition for compressing videos captured by fixed cameras," *Journal of Visual Communication and Image Representation*, vol. 26, no. C, pp. 338–348, 2015.

[16] Beijing Chen, Jianhao Yang, Byeungwoo Jeon, and Xinpeng Zhang, "Kernel quaternion principal component analysis and its application in RGB-D object recognition," *Neurocomputing*, vol. 266, pp. 293 – 303, 2017.

[17] Raghav Subbarao and Peter Meer, "Subspace estimation using projection based M-estimators over grassmann manifolds," in *Computer Vision - ECCV 2006, European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings*, 2006, pp. 301–312.

[18] Nuno Pinho Da Silva and Joao Paulo Costeira, "Subspace segmentation with outliers: A Grassmannian approach to the maximum consensus subspace," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–6.

[19] Anthony Mandow, Jorge L. Martnez, Antonio J. Reina, and Jess Morales, "Fast range-independent spherical subsampling of 3D laser scanner points and data reduction performance evaluation for scene registration," *Pattern Recognition Letters*, vol. 31, no. 11, pp. 1239 – 1250, 2010.

[20] Andreas Geiger, Philip Lenz, and Raquel Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.

[21] P. J. Besl and Neil D. Mckay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.