

Technische Universität München  
TUM School of Computation, Information and Technology

# Robust Solutions for Pose Estimation Problems in Intelligent Transportation Systems

**Xinyi Li**

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung eines

**Doktors der Naturwissenschaften (Dr. rer. nat.)**

genehmigten Dissertation.

**Vorsitz:** Prof. Dr. Francisco Javier Esparza Estau

**Prüfer der Dissertation:**

1. Prof. Dr.-Ing. habil. Alois C. Knoll
2. Prof. Dr. Feihu Zhang

Die Dissertation wurde am 26. 06. 2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 08. 11. 2023 angenommen.





## Abstract

Modern intelligent transportation systems (ITSs) integrate various advanced technologies with transportation infrastructures to significantly improve road safety and traffic efficiency. Pose estimation is an essential geometric perception problem in ITSs, which relies on measurements from various sensors, such as the camera, radar, and light detection and ranging (LiDAR), etc., to accurately estimate the position and orientation of objects within the traffic environment. To meet system safety requirements, pose estimation techniques must be able to effectively handle a large number of outliers, thus ensuring reliable and accurate results.

Branch and bound (BnB) is an algorithm framework for deterministic global optimization, which is highly robust against noise and outliers and can provide globally optimal solutions with theoretical guarantees. Therefore, this thesis aims to explore global and robust solutions based on BnB for two typical categories of pose estimation problems in ITSs, comprising sensor extrinsic calibration and point cloud registration. In conclusion, the main contributions are as follows:

- This thesis proposes a globally optimal and robust extrinsic calibration method for traffic radar. The proposed calibration method is targetless, which can tackle the inconvenience during ITS operation. Besides, a novel BnB-based globally optimal registration method is proposed, which aligns the measurements of traffic radar and GPS in a simultaneous pose and correspondence registration manner. Practical experiments demonstrate that the proposed method is robust to noise and outliers in radar measurements and is able to avoid failure by getting trapped in local optima in the case of large relative angles.
- This thesis proposes an efficient and robust method for the LiDAR registration problem. Specifically, a novel pose decoupling strategy based on residual projections is introduced to effectively decompose the raw problem into three search sub-problems. Compared with existing methods, the proposed approach searches for the optimal solution in the low-dimensional solution domain, thereby improving search efficiency. Subsequently, a novel upper bound based on interval stabbing is derived, and a BnB-based deterministic search algorithm is proposed to address these sub-problems. Extensive experiments demonstrate that the proposed method outperforms state-of-the-art methods in terms of efficiency while simultaneously ensuring robustness.
- This thesis proposes an efficient and robust method for the LiDAR registration problem with gravity prior. First, a novel decoupling scheme is proposed, which can decouple the joint transformation into separate translation and rotation with the aid of known

gravity directions. Subsequently, a BnB-based deterministic optimization algorithm for the translation search sub-problem is proposed, followed by an efficient global voting algorithm for the rotation estimation sub-problem. Compared with the state-of-the-art methods, the proposed method is not only more robust against outliers and biases in gravity directions but also more efficient.

- By further exploring the geometric constraints from the perspective of screw theory, this thesis proposes another efficient and robust method for LiDAR registration with gravity prior. Specifically, a novel transformation decoupling strategy is proposed by leveraging screw theory, which can decompose the original registration problem into three sub-problems. Accordingly, a novel and deterministic three-stage search strategy for the decoupled sub-problems is proposed, which contains interval stabbing, BnB, and global voting algorithms. Extensive experiments demonstrate that the proposed method is more efficient and robust than state-of-the-art methods.

Most of the main contents were published in international journals, indicating the originality and reliability of the work.

## Zusammenfassung

Moderne intelligente Transportsysteme (ITS) integrieren verschiedene fortschrittliche Technologien in die Verkehrsinfrastruktur, um die Verkehrssicherheit und Effizienz erheblich zu verbessern. Die Pose-Schätzung ist ein wesentliches geometrisches Wahrnehmungsproblem in ITS, das auf Messungen von Sensoren wie Kamera, Radar und Lichterkennung und Abstandsmessung (LiDAR) basiert, um die Position und Ausrichtung von Objekten innerhalb der Verkehrsumgebung genau zu schätzen. Um den Sicherheitsanforderungen des Systems zu entsprechen, müssen Pose-Schätztechniken in der Lage sein, eine große Anzahl von Ausreißern effektiv zu handhaben, um so zuverlässige und genaue Ergebnisse zu gewährleisten.

"Branch and Bound" (BnB) ist ein Algorithmusrahmen für die deterministische globale Optimierung, der äußerst robust gegenüber Störungen und Ausreißern ist und theoretisch garantierte global optimale Lösungen liefern kann. Daher zielt diese Arbeit darauf ab, globale und robuste Lösungen auf der Grundlage von BnB für zwei typische Kategorien von Problemen bei der Pose-Schätzung in ITS zu erforschen, nämlich die extrinsische Kalibrierung von Sensoren und die Registrierung von Punktwolken. Zusammenfassend sind die Hauptbeiträge wie folgt:

- Diese Arbeit schlägt eine global optimale und robuste extrinsische Kalibrierungsmethode für Verkehrsradare vor. Die vorgeschlagene Kalibrierungsmethode ist ziellos, was die Unannehmlichkeiten während des Betriebs von ITS bewältigen kann. Darüber hinaus wird eine neuartige, auf BnB basierende global optimale Registrierungsmethode vorgeschlagen, die die Messungen von Verkehrsradar und GPS in einer simultanen Pose- und Korrespondenzregistrierungsweise ausrichtet. Praktische Experimente zeigen, dass die vorgeschlagene Methode robust gegenüber Störungen und Ausreißern in Radarmessungen ist und Ausfälle vermeiden kann, indem sie nicht in lokalen Optima bei großen relativen Winkeln stecken bleibt.
- Diese Arbeit schlägt eine effiziente und robuste Methode für zur LiDAR-Registrierung vor. Speziell wird eine innovative Pose-Decoupling-Strategie auf Basis von Residualprojektionen vorgestellt, um das Rohproblem effektiv in drei Teilprobleme aufzuteilen. Im Vergleich zu bestehenden Methoden sucht der vorgeschlagene Ansatz nach der optimalen Lösung im niederdimensionalen Lösungsbereich, wodurch die Sucheeffizienz verbessert wird. Anschließend wird eine neue Obergrenze auf Basis von Intervall-Stubbing abgeleitet, und ein auf Branch-and-Bound basierender deterministischer Suchalgorithmus wird vorgeschlagen, um diese Teilprobleme zu bewältigen. Umfangreiche Experimente zeigen, dass die vorgeschlagene Methode in Bezug auf Effizienz die Methoden auf dem neuesten Stand der Technik übertrifft und gleichzeitig Robustheit gewährleistet.
- Diese Arbeit schlägt eine effiziente und robuste Methode zur LiDAR-Registrierung mit vorgegebener Schwerkraft vor. Zunächst wird ein neuartiges Entkopplungsschema vorgeschlagen, das die gemeinsame Transformation mithilfe bekannter Schwerkrichtungen in sep-

arate Translation und Rotation aufspalten kann. Anschließend wird ein auf Branch-and-Bound basierender deterministischer Optimierungsalgorithmus für das Teilproblem der Translationsuche vorgeschlagen, gefolgt von einem effizienten globalen Abstimmungsalgorithmus für das Teilproblem der Rotationsabschätzung. Im Vergleich zu Methoden auf dem neuesten Stand der Technik ist die vorgeschlagene Methode nicht nur robuster gegenüber Ausreißern und Verzerrungen in den Schwerkraftrichtungen, sondern auch effizienter.

- Durch eine weitere Untersuchung der geometrischen Einschränkungen aus der Perspektive der Schraubentheorie schlägt diese Arbeit eine weitere effiziente und robuste Methode für die LiDAR-Registrierung mit vorgegebener Schwerkraft vor. Speziell wird eine innovative Transformationsentkopplungsstrategie mithilfe der Schraubentheorie vorgeschlagen, die das ursprüngliche Registrierungsproblem in drei Teilprobleme zerlegen kann. Entsprechend wird eine neuartige und deterministische Dreistufige Suchstrategie für die entkoppelten Teilprobleme vorgeschlagen, die Interval Stabbing, BnB und globale Abstimmungsalgorithmen enthält. Umfangreiche Experimente zeigen, dass die vorgeschlagene Methode effizienter und robuster ist als Methoden auf dem neuesten Stand der Technik.

Der Großteil der Hauptinhalte wurde in internationalen Fachzeitschriften veröffentlicht, was die Originalität und Zuverlässigkeit der Arbeit belegt.

## Acknowledgement

It has been a remarkable and unforgettable journey pursuing my Ph.D. degree in the Chair of Robotics, AI, and Real-time Systems at Technical University of Munich, Germany. Throughout this experience, I have had the pleasure of meeting many incredible people. First and foremost, I wish to convey my profound appreciation to my supervisor, Prof. Alois Knoll. I am truly thankful for his acceptance as my Ph.D. supervisor and for creating the best working environment. His substantial support and insightful guidance were instrumental in ensuring the successful execution of my research. Without his support, the completion of the work presented in this dissertation would have been beyond reach.

I would like to express my gratitude to Dr. Yinlong Liu for generously sharing his scientific knowledge and providing valuable suggestions for my research. His intelligence and profound insights into mathematics and geometry have been consistently impressive. I would like to thank my mentor, Prof. Feihu Zhang, for providing me with valuable suggestions and support for this dissertation. Their openness in sharing time, expertise, and resources, along with their consistent support and encouragement, have been invaluable during my doctoral journey. Additionally, I would like to express my appreciation to Prof. Javier Esparza for chairing the examination committee.

I would like to thank the entire team at our Chair of Robotics, AI, and Real-time Systems for our shared time. Special thanks to Amy Bücherl, Ute Lomp, and Delle Janine for their support and help from the beginning of my Ph.D. to graduation. I am also grateful to my colleagues, Hu Cao, Genghang Zhuang, Yingbai Hu, Christian Creß, Walter Zimmer, Leah Strand, Venkatnarayanan Lakshminarasimhan, Edmond Irani Liu, Xiao Wang, Wenjun Liu, Ligu Zhou, Zhuangyi Jiang, and Dr. Zhengshan Bing. I cherish the countless memories we have created together over the years.

Lastly, I want to express my deepest appreciation to my mother, Yanqing Zhang, and my father, Wengang Li. I am filled with gratitude for the unwavering support, encouragement, and love you have provided throughout this challenging yet rewarding experience.



# Contents

<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Overview . . . . .	1
1.2 Problems and Challenges . . . . .	3
1.2.1 Sensor Extrinsic Calibration . . . . .	4
1.2.2 Point Cloud Registration . . . . .	4
1.3 Contributions . . . . .	5
1.4 Outline . . . . .	7
<b>2 Related Work</b>	<b>9</b>
2.1 Sensor Extrinsic Calibration . . . . .	9
2.1.1 Target-based Calibration . . . . .	9
2.1.2 Targetless Calibration . . . . .	10
2.2 Point Cloud Registration . . . . .	12
2.2.1 Correspondence-based Registration . . . . .	12
2.2.2 Correspondence-free Registration . . . . .	14
<b>3 Globally Optimal Robust Radar Calibration</b>	<b>17</b>
3.1 Background . . . . .	17
3.2 Problem Formulation . . . . .	20
3.2.1 ITS Coordinate System . . . . .	20
3.2.2 Measurement Set Registration Problem . . . . .	21
3.3 Method . . . . .	22
3.3.1 Gaussian Mixture Model and Similarity Measures . . . . .	22
3.3.2 BnB and Relaxation Bounds . . . . .	24
3.3.3 GMRBnB Algorithm . . . . .	28
3.4 Experiments . . . . .	29
3.4.1 Experimental Setup and Convergence . . . . .	29
3.4.2 Control Experiments on Synthetic Data . . . . .	30
3.4.3 Calibration Experiments on Real-World Data . . . . .	33
3.5 Summary . . . . .	39

<b>4</b>	<b>Efficient and Robust LiDAR Registration</b>	<b>41</b>
4.1	Background . . . . .	41
4.2	Problem Formulation . . . . .	44
4.2.1	Inlier Set Maximization . . . . .	44
4.2.2	Residual Projections and Pose Decoupling . . . . .	45
4.3	Step-wise Search Strategy Based on Branch and Bound . . . . .	47
4.3.1	Parametrization of Solution Domain . . . . .	47
4.3.2	Interval Stabbing and Bounds . . . . .	49
4.3.3	Simultaneous Pose and Correspondence Registration . . . . .	53
4.4	LiDAR Registration Experiments . . . . .	56
4.4.1	Experimental Setting . . . . .	56
4.4.2	Synthetic Data Experiments . . . . .	58
4.4.3	Real-World Data Experiments . . . . .	62
4.4.4	Sensitivity Analysis . . . . .	69
4.5	Summary . . . . .	70
<b>5</b>	<b>Efficient and Robust LiDAR Registration with Gravity Prior</b>	<b>73</b>
5.1	Background . . . . .	73
5.2	Transformation Decoupling by Gravity Direction . . . . .	75
5.2.1	Problem Formulation . . . . .	77
5.2.2	Decoupled 4-DOF Point Cloud Registration . . . . .	79
5.3	Transformation Decoupling by Screw Theory . . . . .	85
5.3.1	Problem Formulation . . . . .	86
5.3.2	Three-Stage Consensus Maximization Registration . . . . .	90
5.3.3	Simultaneous Pose and Correspondence Registration . . . . .	98
5.4	LiDAR Registration Experiments . . . . .	99
5.4.1	Experiments for the Two-Stage Method . . . . .	100
5.4.2	Experiments for the Three-Stage Method . . . . .	110
5.4.3	Registration of LiDARs in ITS Infrastructure . . . . .	124
5.5	Summary . . . . .	126
<b>6</b>	<b>Conclusion</b>	<b>129</b>
6.1	Summary . . . . .	129
6.2	Primary Contributions . . . . .	130
6.3	Limitations . . . . .	131
6.4	Future Directions . . . . .	131
6.4.1	Sensor Extrinsic Calibration . . . . .	131
6.4.2	Point Cloud Registration . . . . .	132
	<b>Bibliography</b>	<b>135</b>



# List of Figures

1.1	One of the measurement points on the highway A9 from Providentia++ project ( <a href="https://innovation-mobility.com/en/project-providentia/">https://innovation-mobility.com/en/project-providentia/</a> ). Adapted from Fig. 1 in [Li+23d]. . . . .	2
1.2	A typical pose estimation pipeline for point cloud registration problem. Given two point clouds, the pipeline first extracts and matches keypoints to generate putative correspondences (referred to as the <i>matching front-end</i> ). Next, the pipeline estimates the best rotation and translation to align two point clouds by solving a mathematical optimization problem (referred to as the <i>optimization back-end</i> ). The source, target, and aligned point clouds are green, orange, and blue, respectively. . . . .	3
1.3	The framework of this thesis. . . . .	7
3.1	This is the image after the traffic radar measurements being projected onto the camera picture. The measurements (blue points) are sparse and contain biases and misdetections caused by noise. . . . .	19
3.2	The top view of GPS coordinate system. . . . .	20
3.3	The traffic radar coordinate system and its top view. . . . .	21
3.4	The ITS coordinate system related to the extrinsic calibration. . . . .	21
3.5	The representation of measurement sets $\mathcal{X}$ and $\mathcal{Y}$ with GMMs $\mathcal{G}_\mathcal{X}$ and $\mathcal{G}_\mathcal{Y}$ . . . . .	24
3.6	Search domain parametrization for BnB. The rotation domain is the range of rotation angle, i.e., $[\theta^-, \theta^+]$ . The translation is assumed to be a 2D rectangle $[\mathbf{t}^-, \mathbf{t}^+]^2$ . The splitting policy is to divide from the center of each dimension of the domain, so there will be $2^3 = 8$ sub-boxes after one splitting. . . . .	25
3.7	The schematic of the relaxation lower bound function $R_L(\mathbb{B})$ with two cases, and $e_L$ is the abbreviation of the lower bound of the point-to-point residual $e_{L, X_i, Y_j}(\theta, \mathbf{t})$ . . . . .	26
3.8	Convergence curves of the proposed GMRBnB algorithm. . . . .	29
3.9	An example of measurement sets pair before and after calibration. . . . .	30
3.10	Success rate of CPD, GMMReg, ICP, and GMRBnB algorithms without translation. . . . .	31
3.11	Control experiments on synthetic data with different outlier rates. . . . .	32
3.12	Control experiments on synthetic data with different noise levels. . . . .	33
3.13	This is the test car with a high-precision RTK-GPS device (cm-level). . . . .	34

3.14	The initial measurement sets of radar and GPS are $m = 851$ and $n = 637$ , respectively. The calibration results of manual association, GMRBnB, ICP, GMM-Reg, and CPD for Radar RA are compared. . . . .	35
3.15	The initial measurement sets of radar and GPS are $m = 1199$ and $n = 896$ , respectively. The calibration results of manual association, GMRBnB, ICP, GMM-Reg, and CPD for Radar RB are compared. . . . .	35
3.16	The initial measurement sets of radar and GPS are $m = 1122$ and $n = 881$ , respectively. The calibration results of manual association, GMRBnB, ICP, GMM-Reg, and CPD for Radar RC are compared. . . . .	36
3.17	After projecting the detection of Radar RA (blue points) onto the camera image, the performance of the proposed calibration method is compared before and after its application in three different scenarios. . . . .	36
3.18	After projecting the detection of Radar RB (blue points) onto the camera image, the performance of the proposed calibration method is compared before and after its application in three different scenarios. . . . .	37
3.19	After projecting the detection of Radar RC (blue points) onto the camera image, the performance of the proposed calibration method is compared before and after its application in three different scenarios. . . . .	37
4.1	The proposed method can efficiently address the rigid registration problem in different scenarios with high outlier rates or low overlap rates. For the correspondence-based registration problem, the input correspondences are generated by the traditional descriptor FPFH [RBB09] and the learning-based descriptor FCGF [CPK19]. The input point clouds are selected from (a) Bremen dataset [BEN13], (b) ETH dataset [TWS14], (c) KITTI dataset [GLU12], and (d) Bunny dataset [CL96], respectively. The source point cloud is green, the target point cloud is yellow, and the aligned point cloud is blue. Compared with state-of-the-art (SOTA) correspondence-based methods, the proposed method achieves significant performance in terms of robustness and efficiency. Besides, the proposed method also can solve the SPCR problem efficiently and robustly. . . . .	43
4.2	A toy 2D registration example to demonstrate $L_\infty$ residual projection. Specifically, $\{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^3$ is the set of input correspondences, $\mathbf{r}_j$ , $j = X, Y$ , is the transpose of each row of the rotation matrix, and $t_j$ , $j = X, Y$ , is the component of the translation vector. The red line segments represent the projections of the residual on the coordinate axes $X$ and $Y$ , i.e., $ \mathbf{r}_j^T \mathbf{p}_1 + t_j - q_1^j $ , $j = X, Y$ . The inlier constraint for $L_\infty$ residual indicates that $(\mathbf{p}_1, \mathbf{q}_1)$ is an inlier only if both residual projections on the coordinate axes are not larger than the inlier threshold. . . . .	45

4.3	The solution domain before and after <i>exponential mapping</i> , and the pipeline of the proposed BnB algorithm. The original solution domain of the vector $\mathbf{r}$ is a <i>unit sphere</i> in the 3D Euclidean space. The exponential mapping method maps the unit sphere to two identical 2D disks, representing the solution domains of $\mathbf{r}$ and $-\mathbf{r}$ , respectively. We can only branch one 2D-disk domain during each iteration, followed by the calculation of upper and lower bounds for each sub-branch. The proposed BnB algorithm converges until the optimal solution $\mathbf{r}^*$ is found, and the optimal $t^*$ is found by <i>interval stabbing</i> simultaneously. In the visualization results of interval stabbing, the black line segments are the candidate intervals of each correspondence, and the red line segments are the intervals crossed by the blue probe with the <i>max-stabbing number</i> . The probe position is the <i>max-stabbing position</i> . . . . .	47
4.4	The visualization results for the 24-th iteration of a representative SPCR test on synthetic data. The black line segments are the intervals after <i>interval merging</i> , and the red line segments are the intervals crossed by the blue probe with the max-stabbing number. After interval merging, interval stabbing is utilized to calculate the upper bounds. Notably, the blue probe resulting from interval stabbing can only penetrate at most one interval for each point $\mathbf{p}_i$ . The final SPCR upper bound for $\mathbb{S}^2$ is 38. . . . .	52
4.5	Controlled experiments with $N = \{1000, 2000, \dots, 5000\}$ . The results include average rotation errors, average translation errors, and average running times. . . . .	59
4.6	Controlled experiments with the same experimental conditions as TR-DE [Che+22c]. We use the gray rectangular region to approximately represent the results of TR-DE given in [Che+22c]. The results include average rotation errors, average translation errors, and average running times. . . . .	59
4.7	Controlled experiments with $\eta = \{0.1, 0.2, \dots, 0.8\}$ . The results include average rotation errors, average translation errors, and average running times. . . . .	61
4.8	Controlled SPCR experiments with $\rho = \{0.9, 0.8, \dots, 0.4\}$ . The results include rotation errors, translation errors, and average running times. . . . .	61
4.9	Experiment results on the ETH dataset [TWS14] with the FPFH [RBB09] descriptor. The results include rotation errors, translation errors, and running times. . . . .	64
4.10	Registration results of the proposed method on the ETH dataset [TWS14], including four scan pairs: (a) Courtyard1, (b) Facade1, (c) Office1, and (d) Trees1. The aligned source point cloud is blue, and the target point cloud is yellow. . . . .	65
4.11	Experiment results on the Bremen dataset [BEN13] with the FPFH [RBB09] descriptor. The results include rotation errors, translation errors, and running times. . . . .	66
4.12	Our 3D scene reconstruction results on the Bremen dataset [BEN13], where different scans are indicated by different colors. The pair-wise point cloud registration is conducted for all 12 scan pairs. . . . .	67

4.13	Controlled experiments with different inlier thresholds $\lambda = \{1, 2, 3, 4, 5\}\sigma$ and outlier rates $\eta = \{0, 0.2, 0.4, 0.6, 0.8\}$ . The results include rotation errors, translation errors, and average running times. . . . .	70
5.1	Principle of our decoupled 4-DOF point cloud registration method. Specifically, the joint pose can be decoupled into the translation and rotation when the gravity directions are known. Our method comprises two procedures: 1) a 3-DOF translation search by BnB and 2) a 1-DOF rotation estimation by global voting. Since the search space of BnB is only 3-DOF, our method is able to significantly reduce the running time compared to the joint 4-DOF BnB-based method [Cai+19]. The illustration shows one pair of terrestrial LiDAR scans from <i>Arch</i> in the ETH dataset. After applying the fast match pruning (FMP) [Cai+19] to the candidate correspondences set, the running times of our method and the joint 4-DOF BnB-based method are 2.6ms and 635ms, respectively. Please note that the gravity directions used here are assumed as $[0, 0, -1]^T$ . . . . .	76
5.2	Geometric principle of the proposed upper and lower bounds for BnB. The uncertainty cube $\mathcal{C}_t^{p_i}$ is the red cube. The center of the cube is $\mathbf{p}_i + \mathbf{t}_0$ . The light gray sphere is the relaxed circumsphere of the cube. The bound for the uncertain angle $\alpha_i$ is $[\alpha_i^-, \alpha_i^+]$ . The splitting policy is to divide from the center of each dimension of the translation domain, i.e., there will be $2^3 = 8$ sub-branches after one split. . . . .	79
5.3	A 2D illustration of the upper bound in Case 1, i.e., the relaxed circumsphere and the origin do not intersect. . . . .	80
5.4	A 2D illustration of the upper bound in Case 2, i.e., the relaxed circumsphere and the origin intersect. The uncertain angle $\alpha_i$ lies arbitrarily within the interval $[-\pi, \pi]$ . . . . .	80
5.5	Geometric intuition for angle $\theta$ , which is the angle between the projections of two vectors $\mathbf{v}_1 = \mathbf{R}_{\mathbf{v}_p}^{\mathbf{v}_q} \cdot (\mathbf{p}_i + \mathbf{t})$ and $\mathbf{v}_2 = \mathbf{q}_i$ . The projections $\mathbf{v}'_1$ and $\mathbf{v}'_2$ are on the plane with $\mathbf{v}_q$ as the normal. . . . .	83
5.6	Schematic of Chasles' Theorem and screw transformation. Specifically, the Euclidean transformation of point $\mathbf{p}_i$ , i.e., $\mathbf{R}\mathbf{p}_i + \mathbf{t}$ can be represented by a screw rotation $\mathcal{S}(\mathbf{p}_i)$ combined with a screw translation $\mathbf{t}_{\parallel}$ . The screw axis is defined by a unique point $\mathbf{C}_i$ in the plane $\Omega_i$ and the rotation axis $\mathbf{r}$ . . . . .	87
5.7	Calculation flow of the proposed three-stage method for correspondence-based registration. The source, target, and aligned point clouds are green, orange, and blue, respectively. In the initial correspondences set, green and red line segments represent inliers and outliers, respectively (only a subset of correspondences is shown for visual clarity). For the interval stabbing part, the candidate intervals are represented by the black line segments, while the intervals crossed by the max-stabbing probe are depicted as the red line segments. . . . .	89

5.8	The geometrical interpretation of the coordinate conversion by $\mathbf{R}_{v_q}^{e_z}$ and the plane projection to $\Omega_0$ . . . . .	92
5.9	The geometrical interpretation of the 2D rigid transformation and the 2D screw rotation after plane projection. . . . .	94
5.10	Exponential mapping of the solution domain and the geometrical interpretation of Lemma 5.1. . . . .	96
5.11	Calculation flow of the proposed three-stage method for correspondence-free registration. The bolded points in the point cloud represent the points after downsampling. For the interval merging and stabbing part, the merged intervals are represented by the black line segments, while the intervals crossed by the max-stabbing probe are depicted as the red line segments. . . . .	99
5.12	Efficient translation search by the proposed BnB and global voting for rotation estimation. . . . .	101
5.13	Visualization of the synthetic point cloud data. Each subplot (a)-(d) contains two figures (before and after registration with our method) and has different parameter settings, as indicated in the subheadings. The source point cloud is shown in red, and the target point cloud is shown in green. The rotation error $e_{\text{rot}}(^{\circ})$ and translation error $e_{\text{trans}}(\text{m})$ for each example are: (a) 0.024 and 0.144, (b) 0.228 and 0.787, (c) 0.323 and 0.953, (d) 0.504 and 1.326, respectively. It is evident that the rotation and translation errors increase significantly as the outlier rate and noise level increase. . . . .	102
5.14	Controlled experiments with normal ( $\sigma = 1.0$ ) and high ( $\sigma = 1.5, 2.0$ ) noise levels; in each group, the outlier rate is $\eta = \{0.1, 0.2, \dots, 0.8\}$ . The results include rotation error, translation error, and median time. . . . .	103
5.15	Controlled experiments with different numbers of correspondences ( $N = \{500, 1000, \dots, 10000\}$ ). The outlier rate is $\eta = 0.5$ and the noise level is $\sigma = 1.0$ . . . . .	105
5.16	Controlled experiments with different biased angles ( $0.1^{\circ}, 0.3^{\circ}, 0.5^{\circ}$ ) in gravity directions and different outlier rates $\eta = \{0.1, 0.2, \dots, 0.8\}$ . . . . .	105
5.17	Challenging SPCR experiments with different outlier rates $\eta = \{0.1, 0.2, \dots, 0.6\}$ in the Bunny dataset. . . . .	105
5.18	Rotation error and translation error of all registration methods on the ETH dataset. Subfigures (a)-(e) show the registration results of each scan pair for <i>Arch</i> , <i>Courtyard</i> , <i>Facade</i> , <i>Office</i> , and <i>Trees</i> . . . . .	108
5.19	Controlled experiments on the outlier rate. (a) Median running time, (b) Rotation error, and (c) Translation error. . . . .	112
5.20	Controlled experiments on the number of correspondences. (a) Median running time, (b) Rotation error, and (c) Translation error. . . . .	114
5.21	Controlled experiments on the gravity direction noise. (a) Median running time, (b) Rotation error, and (c) Translation error. . . . .	114
5.22	Controlled SPCR experiments on the overlap rate. (a) Median running time, (b) Rotation error, and (c) Translation error. . . . .	114

5.23	Examples of scan pairs and initial correspondences for each scene from ETH dataset. The number of correspondences and outlier rate are denoted by $N$ and $\eta$ , respectively. . . . .	115
5.24	Running times, rotation errors, and translation errors of all registration methods on the ETH dataset. (a)-(e) show the registration results for <i>Arch</i> , <i>Courtyard</i> , <i>Facade</i> , <i>Office</i> , <i>Trees</i> . . . . .	116
5.25	(a) Running time, (b) Rotation error, and (c) Translation error of all registration methods on the WHU-TLS dataset. . . . .	117
5.26	Qualitative and quantitative registration results for scan pair <i>Railway</i> <sub>1</sub> on the WHU-TLS dataset. . . . .	119
5.27	Qualitative correspondence-free registration results on the Stanford and UWA datasets. (a) <i>Stanford-Armadillo</i> , (b) <i>Stanford-Dragon</i> , (c) <i>U3M-Chef</i> , (d) <i>U3M-Chicken</i> , (e) <i>U3M-Parasaurolophus</i> , (f) <i>U3M-T-rex</i> . The source, target, and aligned point clouds are green, orange, and blue. . . . .	123
5.28	One of the measurement points of the Providentia++ project on the A9 highway ( <a href="https://innovation-mobility.com/en/project-providentia/a9-dataset/">https://innovation-mobility.com/en/project-providentia/a9-dataset/</a> ). The A9 dataset [Cre+22] is obtained from roadside sensors, including multiple LiDARs, cameras, and radars. . . . .	125
5.29	Qualitative results for scan pair P3 on the A9 dataset. (a) Initial, (b) GROR, (c) FMP+Ours1, (d) Ours2. . . . .	126

# List of Tables

3.1	Comparison of rotation error and translation error of all algorithms for three radars . . . . .	38
4.1	Controlled experiments with the extremely high number of correspondences. The results include average rotation error ( $^{\circ}$ )   average translation error (m)   average running time (s). Bolded and underlined fonts indicate the first two best values. . . . .	58
4.2	Detailed information about the ETH Dataset. . . . .	63
4.3	Detailed information about the Bremen Dataset. . . . .	65
4.4	Experiment results on the KITTI dataset[GLU12] with FCGF [CPK19] descriptors. Bolded and underlined fonts indicate the first two best values. . . . .	68
4.5	Orthogonality and determinant evaluation results on the KITTI dataset. The evaluation metrics comprise $O_{XY}$ , defined as the inner product of $\mathbf{r}_X^*$ and $\mathbf{r}_Y^*$ , alongside $O_{XZ}$ , $O_{YZ}$ , and the determinant $Det$ of $[\mathbf{r}_X^*, \mathbf{r}_Y^*, \mathbf{r}_Z^*]^T$ . The mean and standard deviation of these metrics are given. . . . .	69
5.1	Success rates $r_1$ and $r_2$ (%) with successful cases satisfying Threshold 1 ( $e_{rot} \leq 0.5^{\circ}, e_{trans} \leq 0.2m$ ) and Threshold 2 ( $e_{rot} \leq 5^{\circ}, e_{trans} \leq 2m$ ), respectively, on the ETH dataset . . . . .	104
5.2	The left part is detailed information about the ETH dataset. The right part is the registration running time (ms) for each pair of scans on the ETH dataset. .	106
5.3	Controlled experiments on the number of correspondences in the extreme case. The results include median running time (s)   success rate (%) with cases satisfying $RE \leq 1^{\circ}$ and $TE \leq 0.01m$ . . . . .	113
5.4	Detailed information about the ETH dataset . . . . .	120
5.5	Detailed information about the WHU-TLS dataset . . . . .	121
5.6	Quantitative registration results on real-world datasets. The results include average running time (s)   average rotation error ( $^{\circ}$ )   average translation error (m). Bolded fonts indicate the best results. . . . .	121
5.7	Detailed information about the real-world datasets in SPCR experiments. . . .	122
5.8	Quantitative correspondence-free registration results on real-world datasets, including running time (s)   rotation error ( $^{\circ}$ )   translation error (m). Bolded fonts indicate the best results. . . . .	122
5.9	Detailed information about the A9 dataset . . . . .	125

5.10 Quantitative registration results on the A9 dataset. The results include running time (s) | rotation error (°) | translation error (m). Bolded fonts indicate the best results. . . . . 126



# 1

## Introduction

### 1.1 Overview

Modern intelligent transportation systems (ITSs) integrate various advanced information, communication, and sensing technologies with transportation infrastructures to significantly improve road safety and traffic efficiency. For instance, ITSs can provide autonomous vehicles, as well as traditional vehicles, with extra detailed information about traffic participants and the overall traffic situation, thus extending their perception field [Zha+11; Men+17; Krä+19; Ye+22]. Providing road users with this supplementary information enhances their spatial understanding of the surrounding environment, allowing for safer and more proactive maneuver planning. Moreover, ITSs with these capabilities can provide a variety of services to further support decision-making, such as emergency vehicle notification, variable speed limits, and collision warning. As an important frontline module in ITSs, the perception system commonly utilizes a diverse range of sensors to acquire live traffic data in various modalities, including cameras, event cameras, radar, and light detection and ranging (LiDAR), etc. The typical infrastructures of ITS equipped with multiple sensors are given in Fig. 1.1.

Pose estimation is an essential geometric perception problem in ITSs, which leverages sensor measurements to accurately estimate the position and orientation of objects within the traffic environment. Specifically, it is a task of estimating the relative transformation, including rotation and translation, between the object of interest and the reference coordinate system from multiple kinds of sensor measurements (e.g. visual, inertial, and laser-scan data) [Wu+20; Li+23g]. Nowadays, pose estimation problems have a wide range of applications in ITS-related perception tasks, including vehicle localization and mapping [Li+23b; Zhu+22; Li+23a], object detection [Guo+14; Che+23a], camera pose estimation [Che+22b; LCK23], sensor extrinsic calibration [Bel+22; Jia+23b], and point cloud registration [Zha+23b; Yan+23; Che+23b]. It is worth noting that numerous high-level planning and decision-making tasks within ITS heavily depend on the information delivered by the perception system. If inaccurate pose estimation results are provided, it could prompt inappropriate actions taken by traffic participants, potentially leading to severe failures. Hence, it is crucial for pose estimation algorithms to deliver satisfactory solutions.

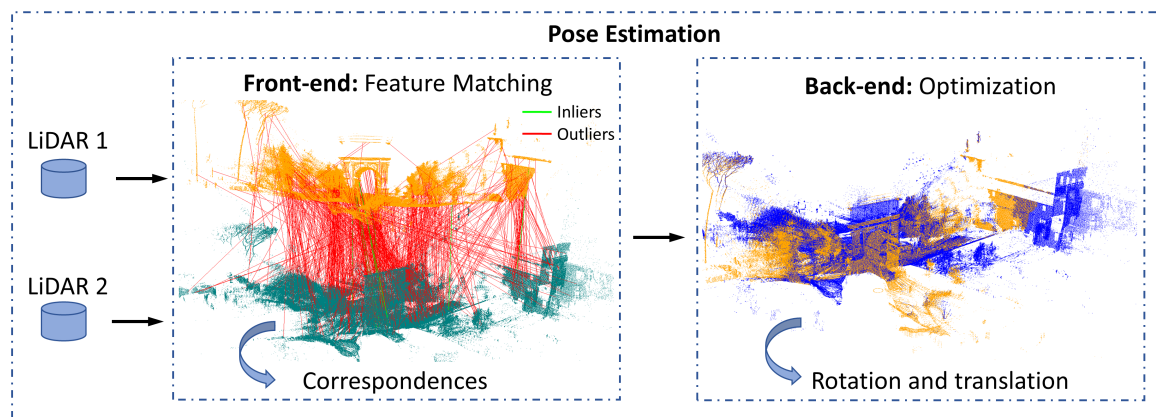


**Figure 1.1:** One of the measurement points on the highway A9 from Providentia++ project (<https://innovation-mobility.com/en/project-providentia/>). Adapted from Fig. 1 in [Li+23d].

A typical pose estimation pipeline including a *matching front-end* that extracts and matches relevant features from original sensor measurements, and an *optimization back-end* that estimates the relative transformation from the given putative feature matches (i.e., correspondences) [YC22; YP23]. However, in practical applications, due to various imperfections and uncertainties (e.g., sensor failures, noisy measurements, and incorrect matching), the presence of *outliers* in feature correspondences is inevitable in the front-end. It is widely recognized that even a single outlier can significantly distort the estimation results [TAC19; Ant+21]. Therefore, the existence of outliers significantly elevates the safety risk in ITS. To meet safety demands, it is imperative for the pose estimation technology to incorporate an *outlier-robust* back-end capable of effectively handling substantial amounts of outliers.

To further explain the importance of a robust back-end in pose estimation technologies, the 3D point cloud registration task is taken as an example. As shown in Fig. 1.2, the front-end first establishes the 3D feature correspondences between the source and target point clouds by 3D descriptors, e.g., fast point feature histograms (FPFH) [RBB09]. When the correspondences are all *inliers*, the back-end can provide an elegant *closed-form solution* [Hor87; AHB87], although the non-convexity of the special orthogonal group  $SO(3)$ . Nonetheless, it is almost impossible to generate the feature correspondences without outliers in practice, and a significant number of outliers is quite common (often  $> 95\%$  [BC17; Li+23b]), such as shown in Fig. 1.2. Therefore, obtaining robust solutions for pose estimation problems in the presence of outliers is crucial. Random sample consensus (RANSAC) [FB81] is one of the most popular methods to suppress outlier correspondences. Unfortunately, due to its randomized nature, RANSAC is non-deterministic and only generates satisfactory results with a certain probability [Li09; Le+19]. In other words, RANSAC probably returns an unsatisfactory solution without notice. In practical applications, these incorrect results may cause misjudgment of the ITS perception system, leading consequent modules in the system to take risky actions.

To obtain highly reliable and robust solutions, an increasing number of *deterministic*



**Figure 1.2:** A typical pose estimation pipeline for point cloud registration problem. Given two point clouds, the pipeline first extracts and matches keypoints to generate putative correspondences (referred to as the *matching front-end*). Next, the pipeline estimates the best rotation and translation to align two point clouds by solving a mathematical optimization problem (referred to as the *optimization back-end*). The source, target, and aligned point clouds are green, orange, and blue, respectively.

*global optimization* algorithms have been proposed to address pose estimation problems recently [Li09; Flo13; Chi+17; CCK19; Le+19; Fan+21]. Most of these methods employ the branch and bound (BnB) optimization algorithm [HK09; Li09; ZSO11; BSP12; Yan+16; LZY16]. The BnB algorithm is a globally optimal optimization technique capable of addressing many non-convex and NP-hard problems [Sch11]. From the optimization aspect, the BnB algorithm systematically searches the entire parameter space to seek the global optimum so that it can avoid local optima. The BnB algorithm has two significant advantages: 1) it can provide the globally optimal solution with theoretical guarantees, and 2) it is independent of initial parameters. Particularly, employing BnB necessitates the development of specific lower and upper bound functions tailored to different target problems. Therefore, this thesis delves into developing robust and accurate back-end optimization methods based on BnB to solve typical pose estimation problems.

In a nutshell, the pose estimation problem has a wide range of applications in ITSs, especially as a core component of many geometric perception tasks, where a robust back-end is extremely important. However, the existence of outliers, almost inevitable in practical applications, will significantly affect the perception performance and increase the ITS safety risk. Unfortunately, conventional pose estimation algorithms commonly cannot guarantee to provide optimal solutions from outlier-contaminated data. Therefore, this thesis presents an investigation of BnB-based robust solutions for typical pose estimation problems in ITSs so as to improve perception performance and system safety.

## 1.2 Problems and Challenges

Pose estimation problems refer to the general task of estimating the transformation of an object from a given reference pose. This thesis mainly focuses on two typical categories of pose estimation problems in ITSs, including sensor extrinsic calibration and point cloud

registration.

### 1.2.1 Sensor Extrinsic Calibration

The ITS commonly utilizes heterogeneous sensors, such as cameras, LiDAR, and radar, to acquire information on different modalities and significantly expand the perception range. These sensors usually have different installation locations and different fields of view. Thus, sensor extrinsic calibration is necessary for multi-sensor spatial fusion. The purpose of sensor extrinsic calibration is to determine how measurements from different sensors can be transformed into a reference coordinate system and build a consistent environment model. The transformation commonly contains a rotation and a translation.

According to the pipeline in Fig. 1.2, the front-end of sensor extrinsic calibration is building feature correspondences. The existing calibration methods are classified as *target-based* [PMP19; TPH20; DKG21; Bel+22] and *targetless* [Zuñ+19; Yua+21; Che+22a; Sun+22a] depending on the way of extracting features (reviewed in Chapter 2). Then, the back-end estimates the transformation parameters by minimizing the reprojection error between the feature correspondences.

Typically, target-based methods rely on prepared artificial targets, such as checkerboards, retro-reflective landmarks, and custom-built calibration patterns. These methods can generate reliable and high-quality feature correspondences based on the specialized targets, thereby obtaining high-accuracy calibration results. However, target-based methods are laborious and time-consuming for practical scenarios because sometimes the installation of specialized targets is infeasible, such as during ITS operation. Targetless calibration methods commonly build feature correspondences from the environment structure, such as lines, edges, and planar regions. However, these methods suffer from the problem of low-quality feature correspondences since environmental features are unstable and sensors are susceptible to noise, leading to outlier correspondences. In summary, the challenges of sensor extrinsic calibration mainly arise from the following three aspects: 1) feature extraction from noisy measurements, 2) feature matching between multi-modal sensors or unavailable feature matching, and 3) robust pose estimation.

### 1.2.2 Point Cloud Registration

Point cloud refers to a set of 3D points acquired by LiDAR or RGB-D cameras. Given source and target point clouds in different coordinate systems, point cloud registration aims to estimate the 6 degrees of freedom (DOF) transformation to align the two point clouds best. The 6-DOF transformation includes both 3-DOF rotation and 3-DOF translation. Besides, point cloud registration is also known as scan matching or point set registration. On the other hand, autonomous vehicles in ITSs are commonly equipped with inertial measurement units (IMUs), which can provide high-precision gravity directions. Thus, researchers leverage prior information from external sensors, such as IMUs, to assist point cloud registration. With the

aid of gravity directions, the relative rotation is reduced to 1 degree of freedom (DOF). Thus, the original 6-DOF point cloud registration problem is reduced to a 4-DOF problem.

For both point cloud registration problems, the front-end commonly generates feature correspondences by 3D handcrafted or learning-based descriptors from original point clouds. Subsequently, given the putative correspondences, the optimization back-end estimates the relative transformation. However, in cases where generating correspondences is intractable, such as when the point clouds are not sampled densely from smooth surfaces, the registration problem can be tackled only through the optimization back-end, employing a simultaneous pose and correspondence registration paradigm. Therefore, the current registration methods are classified as *correspondence-based* [YSC20; Yan+23; Zha+23b; Che+23b] and *correspondence-free* [BM92; BS03; Yan+16; ZYD21] (reviewed in Chapter 2). Due to the limited performance of existing 3D feature matching methods and the non-convexity of the registration problem, the challenges of point cloud registration include: 1) establishing reliable correspondences, 2) robustness against outliers, and 3) computational complexity.

### 1.3 Contributions

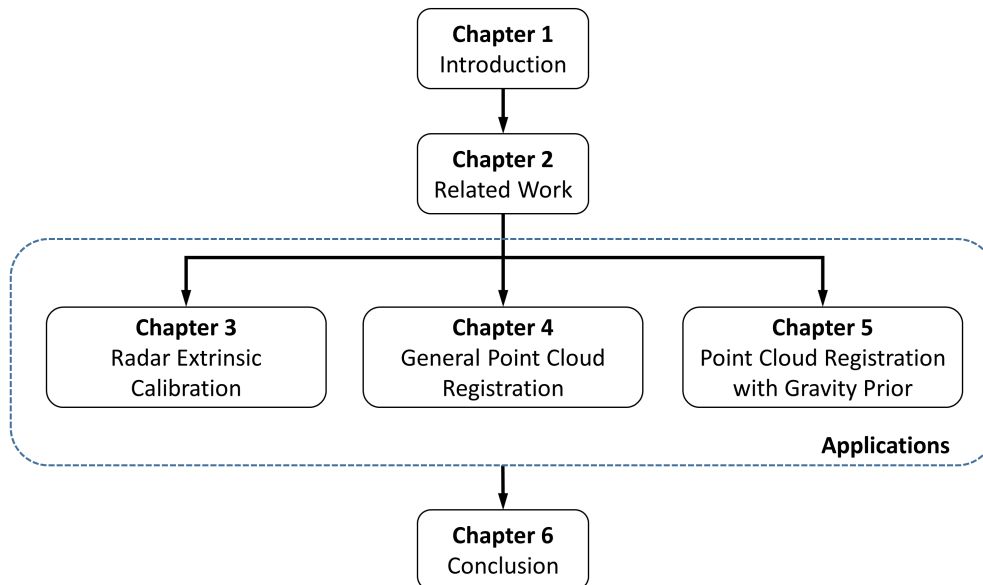
This thesis adopts the deterministic and global BnB algorithm to address the above-mentioned challenges of typical pose estimation problems in ITSs, including sensor extrinsic calibration and point cloud registration. By utilizing the BnB algorithm, the back-end of the pose estimation technique can suppress heavy outlier correspondences and obtain robust solutions. Specifically, the main contributions are concluded as follows:

- This thesis proposes a targetless radar extrinsic calibration method based on GPS to overcome the inconvenience during ITS operation, since the installation of a specialized calibration object on the highway is impractical and dangerous. The high-precision GPS device installed on the moving vehicle can provide traffic radar with accurate positioning information of the detection target. Furthermore, in the back-end of extrinsic calibration, a globally optimal registration method is proposed, which is robust to noise and outliers in radar measurements. Concretely, a robust objective function is first constructed by utilizing the Gaussian mixture model (GMM). Then, novel relaxation bounds are derived for this objective, and a BnB-based algorithm is proposed, which overcomes the susceptibility to local minima and the dependence on the initialization of traditional methods. The proposed algorithm can align the measurements of traffic radar and GPS in a simultaneous pose and correspondence registration (i.e., correspondence-free registration) manner. Compared with existing methods, extensive experiments on synthetic and real-world data demonstrate that the proposed method is not only globally optimal but also more accurate and robust.
- This thesis proposes an efficient and robust method for the general 6-DOF point cloud registration problem. Current 3D feature matching approaches in the front-end commonly generate an overwhelming number of outlier correspondences, which makes

the outlier-robust back-end indispensable. Nonetheless, BnB-based methods are time-consuming to search the entire 6-dimensional parameter space simultaneously since their computational complexity is exponential to the dimensionality of the solution domain in the worst-case. In order to enhance algorithm efficiency, existing works attempt to decouple the 6-DOF original problem into two 3-DOF sub-problems, thereby reducing the dimensionality of the parameter space. In contrast, a novel pose decoupling strategy based on residual projections is introduced in this thesis, which effectively decomposes the raw problem into three search sub-problems. Subsequently, a novel BnB-based deterministic search method is proposed to solve these sub-problems, and the relevant bound functions based on interval stabbing are derived. Moreover, the proposed upper bound is modified by interval merging to solve the registration problem without correspondence. Extensive experiments conducted on synthetic and real-world datasets demonstrate that the proposed method outperforms state-of-the-art methods in terms of efficiency while simultaneously ensuring robustness.

- This thesis proposes an efficient and robust method for the point cloud registration problem with gravity prior. The key idea is to speed up BnB-based methods by decoupling the joint pose into individual translation and rotation with the aid of known gravity directions. This effectively reduces the search domain to 3-DOF+1-DOF, thereby enhancing algorithm efficiency. Specifically, a novel BnB-based consensus maximization method for 3-DOF translation searching is proposed, and the specific lower and upper bound functions are derived. Then, an efficient global voting method to estimate the rotation with 1-DOF is proposed. Extensive experiments on both synthetic and real-world datasets are conducted to demonstrate the superiority of the proposed method. The experimental results show that 1) the proposed method is more robust against outliers and noise than several existing methods and far faster than the existing BnB-based 4-DOF method by almost an order of magnitude, 2) the proposed method is robust against the biases in gravity directions, such that the general error of IMUs is acceptable, and 3) thanks to its significant robustness, the proposed method can solve the challenging correspondence-free registration problem under the all-to-all correspondence assumption.
- By further exploring the geometric properties and constraints, this thesis proposes another efficient and robust method for point cloud registration with gravity prior. Firstly, the registration problem is reformulated by screw theory, and a novel transformation decoupling strategy is proposed accordingly. This strategy is to decompose the 4-DOF original problem into three sub-problems with 1-DOF, 2-DOF, and 1-DOF, respectively, thereby enhancing the algorithm efficiency. Specifically, the first 1-DOF represents the translation along the rotation axis, and an interval stabbing-based method is proposed to estimate it. The second 2-DOF represents the pole which is an auxiliary variable in screw theory. It is reformulated as a linear model fitting problem, and a BnB-based algorithm is utilized to solve it. The last 1-DOF represents the rotation angle, and a global voting method is proposed for its estimation. The proposed method sequen-





**Figure 1.3:** The framework of this thesis.

tially solves three consensus maximization sub-problems to achieve efficient and deterministic registration. In particular, the proposed method can be extended to solve the correspondence-free registration problem without the all-to-all correspondence assumption. Extensive experiments on synthetic and real-world datasets demonstrate that the proposed method is more efficient and robust than state-of-the-art methods.

Most of the contents provided in this thesis are either previously published or under submission for international journals.

## 1.4 Outline

As shown in Fig. 1.3, this thesis is structured as follows:

- Chapter 1 first gives an overview of the pose estimation problems in ITSs and introduces the importance of robust solutions in practical applications. Furthermore, it states two typical categories of pose estimation problems in ITSs and their challenges. Finally, the main contributions are summarized.
- Chapter 2 extensively reviews the related work of sensor extrinsic calibration and point cloud registration.
- Chapter 3 presents a globally optimal robust radar extrinsic calibration method and provides rigorous theoretical derivations and extensive experiment results.

*The main contents of this chapter have been published in IEEE Transactions on Intelligent Transportation Systems [Li+23d].*

- Chapter 4 proposes an efficient and robust method for the point cloud registration problem and provides rigorous theoretical derivations and extensive experiment results. *The main contents of this chapter have been submitted for possible publication in IEEE. The associated arXiv version is [Li+23c].*

- Chapter 5 proposes two novel algorithms for efficient and robust point cloud registration with gravity prior. The first algorithm decouples the transformation by known gravity directions. By further exploring the geometric constraints, the second algorithm decouples the transformation by screw theory. Rigorous theoretical derivations and extensive experiment results are provided.

*The main contents of Section 5.2 have been published in ISPRS Journal of Photogrammetry and Remote Sensing [Li+23e]. The main contents of Section 5.3 have been submitted for possible publication in IEEE. The associated arXiv version is [Li+23f]*

- Chapter 6 comprehensively summarizes the contributions and discusses limitations and future directions.



# 2

## Related Work

*This chapter provides a comprehensive discussion and comparison of the existing work, focusing on sensor extrinsic calibration and point cloud registration domains. Additionally, the strengths and limitations of current methods are summarized and analyzed.*

### 2.1 Sensor Extrinsic Calibration

In the past few decades, researchers have conducted a wide range of studies on sensor extrinsic calibration to improve the perception accuracy of autonomous vehicles and ITSs. Generally, one of the requirements for the front-end of extrinsic calibration is the reference feature, which is extracted from a dedicated target or environment structure. This requirement divides the sensor extrinsic calibration methods into two categories, *target-based* and *targetless* methods.

#### 2.1.1 Target-based Calibration

Target-based methods calibrate the extrinsic parameters of the sensor by using specialized objects that are easily detected, such as checkerboards [Gei+12; Xie+18; ZLK18; Mis+20; Jia+23b], spherical targets [KKL18; KK20; TPH20], reflective landmarks [El +15; PMP19; Lee+20; Olu+21], and custom-built patterns [DKG21; Bel+22]. Compared with targetless methods, target-based methods are commonly more accurate due to the reliable feature correspondences.

Among the various specialized objects, the checkerboard stands out as one of the most commonly used. For example, Geiger *et al.* [Gei+12] proposed a toolbox for the camera-to-LiDAR calibration based on the checkerboard. The well-known iterative closest point (ICP) algorithm is employed for registration in the back-end. Zhou *et al.* [ZLK18] also presented an extrinsic calibration method for LiDAR and camera using the checkerboard. In the front-end, the line and plane correspondences are generated to reduce the number of

required checkerboard poses. Then the parameters are estimated at the back-end using the Levenberg-Marquardt (LM) algorithm. Jiao *et al.* [Jia+23b] proposed a checkerboard-based and globally optimal extrinsic calibration method for LiDAR, camera, and event camera. They introduced a general solver based on Gröbner-basis to globally solve the optimization problems in the back-end. Furthermore, Kümmerle *et al.* [KKL18] employed the spherical target to calibrate the camera and LiDAR. RANSAC is used to achieve outlier-robust pose estimation. Tóth *et al.* [TPH20] established feature correspondences about sphere centers between the camera and LiDAR. Then, they applied the singular value decomposition (SVD) method to obtain the closed-form solution of the relative transformation.

Reflective landmarks are commonly used for millimeter-wave (MMW) radar extrinsic calibration. Natour *et al.* [El +15] established the radar-to-image correspondences by a dedicated target with high reflectivity to implement the calibration between radar and camera. They employed the LM algorithm to determine the relative transformation. Chou *et al.* [CYS17] introduced a mirror-assisted combined calibration object and proposed an extrinsic calibration method for the ground penetrating radar and camera. The LM algorithm is also employed in the back-end to minimize the reprojection error. Peršić *et al.* [PMP19] designed a triangular-shaped corner reflector for radar, LiDAR, and camera calibration. They used the LM algorithm to minimize the reprojection error, followed by radar cross-section (RCS) enhanced optimization. Lee *et al.* [Lee+20] implemented the spatial and temporal calibration for radar and LiDAR by the RCS measurements from a triangular board. In addition, Olutomilayo *et al.* [Olu+21] proposed a radar extrinsic calibration method based on corner reflector targets with the known pose. They optimized the extrinsic parameters by using the Kabsch algorithm. Domhof *et al.* [DKG21] proposed a novel extrinsic calibration tool for radar, camera, and LiDAR by using a specialized target and utilized three joint optimization configurations to perform both relative and absolute calibration. Beltrán *et al.* [Bel+22] presented a custom-built calibration board for LiDAR and camera calibration. The SVD method is used to estimate the optimal relative transformation. The advantage of target-based methods is that they are accurate and easy to implement. However, target-based methods are impractical in ITS-related applications since the installation of dedicated targets on the road is dangerous and infeasible.

### 2.1.2 Targetless Calibration

Different from the target-based methods, the targetless methods commonly extract feature correspondences from the environment structure, such as points [GRG09; ZZM15], lines [MBZ13; BJX20; Zha+21], edges [LT13; KD20; Yua+21], and objects [YJC21; Sun+22b; Sun+22a; Lia+23]. In [MBZ13], authors extracted and matched line features to determine the extrinsic parameters between the camera and LiDAR. Similarly, Zhang *et al.* [Zha+21] proposed a line-based extrinsic calibration method for LiDAR and camera. Then, they presented an adaptive optimization approach to estimate the relative transformation. In [LT13], the 3D edge features were extracted from the point cloud by depth discontinuities and then projected onto the 2D image. They addressed the calibration problem by minimizing the reprojection

error. In [KD20], authors proposed a camera-LiDAR calibration method that minimizes the misalignment of edges by utilizing the gradient descent algorithm. In [Yua+21], authors presented an extrinsic calibration method by aligning edge features in the point cloud and image. They introduced a maximal likelihood optimization method to estimate the extrinsic parameters. Furthermore, Yoon *et al.* [YJC21] employed semantic objects to build feature correspondences and then calibrated extrinsic parameters for the camera and LiDAR. Similarly, Liao *et al.* [Lia+23] extracted semantic features from both point clouds and images, followed by estimating extrinsic parameters. Sun *et al.* [Sun+22b] proposed an attention-based object-level matching network to build 2D-3D correspondences and then presented a particle swarm optimization algorithm to estimate the relative transformation. In [Sun+22a], authors extracted and associated objects by the instance segmentation method and adopted the differential evolution (DE) algorithm to optimize the cost function. In addition to these appearance-based approaches, motion-based methods have also been introduced in [IOI18; Wal+19; Zuñ+19; Yuw+20; LZ21; Jia+21; LYZ22; Das+22; Che+22a], which utilize the motion of sensors as auxiliary cues. These methods commonly formulate extrinsic calibration as a hand-eye calibration problem, where the extrinsic parameters are estimated using the motions of sensors. Nonetheless, these methods typically require sufficient motions to extract co-visible features. Moreover, their accuracy is highly susceptible to the cumulative drifts of the estimated motion.

Due to the sparseness and noisiness of MMW radar measurements, radar can not provide any descriptive features such as edges and corners from the environment structure. Therefore, researchers have utilized additional prior information or a moving platform to associate targets. For example, Schöller *et al.* [Sch+19] proposed a data-driven targetless radar calibration approach based on deep learning to estimate the relative rotation angle, which can avoid the target association problem between camera and radar measurements. However, this approach only achieves the rotational calibration and ignores the translational calibration. Izquierdo *et al.* [Izq+18] proposed a targetless calibration method based on the high-definition (HD) map for multiple radars onboard a vehicle. The reference features are the static objects with specific categories and high RCS, such as street lights and traffic signs. However, they assumed that the vehicle could be localized on the HD map by using GNSS. Similarly, Heng *et al.* [Hen20] proposed a targetless calibration method for multiple 3D LiDARs and 3D radars mounted on a vehicle. They first utilized the LiDAR data to build a 3D map of the environment and then registered the radar scans with the 3D map. They also assumed that known vehicle poses are provided by a GNSS system. Wise *et al.* [Wis+21] utilized continuous-time velocity measurements to implement the extrinsic calibration of onboard 3D radar instead of 2D (planar) radar. This study is based on the ego-motion estimation of a moving vehicle. Du *et al.* [Du+21] introduced a spatio-temporal calibration method for roadside cameras and radar. They proposed an object-matching method based on multiple virtual detection lines to generate object-level correspondences. The objects are extracted from the camera and radar measurements by multi-object tracking methods. However, due to the limited performance of multi-object tracking methods, this calibration method suffers from low accuracy. In general, all the aforementioned targetless methods adhere to

the principle of feature matching, followed by parameter optimization. However, they still encounter problems related to feature extraction and matching, which often result in outlier-contaminated correspondences and thus affect the accuracy of parameter optimization.

## 2.2 Point Cloud Registration

Despite decades of research, the rigid point cloud registration is still an active and challenging problem since it has the chicken-and-egg property [LH07]. Specifically, the registration problem comprises two mutually interlocked sub-problems: pose and correspondence estimations. If one sub-problem is solved, another sub-problem will be solved accordingly. Commonly, existing registration methods are classified into two categories based on the requirement of correspondence or not, which are *correspondence-based registration* and *simultaneous pose and correspondence registration* (a.k.a. *correspondence-free registration*).

### 2.2.1 Correspondence-based Registration

Correspondence-based registration is the 3D extension of image matching, which comprises two major steps: 1) matching correspondences between point clouds by 3D keypoints and feature descriptors, and 2) estimating the transformation from the candidate correspondences. In the matching process, 3D handcrafted or learning-based keypoints (e.g., ISS [Zho09], MeshDoG [Zah+09], Harris3D [SB11], KeypointNet [Suw+18], 3DFeatNet [YL18], USIP [LL19], and RSKDD-Net [Lu+20]) are first extracted from point clouds. Then keypoints are encoded to high-dimensional feature vectors by 3D feature descriptors [TSD13; Guo+16]. Compared to hand-crafted 3D descriptors (e.g., FPFH [RBB09], SHOT [TSD10; STD14], RoPS [Guo+13a], and TriSI [Guo+13b]), the learning-based descriptors (e.g., 3DMatch [Zen+17], 3DSmoothNet [Goj+19], FCGF [CPK19], D3Feat [Bai+20], SpinNet [Ao+21], Predator [Hua+21], and YOHO [Wan+22]) exhibit outstanding precision and have achieved more attention in recent years. Finally, the putative correspondences between point clouds are established by computing pairwise similarity, such as utilizing the nearest neighbor matcher [Low04]. Recently, learning-based matchers (e.g., CoFiNet [Yu+21], GeoTransformer [Qin+22], and RoITr [Yu+23]) are also investigated to improve correspondence quality, which excludes keypoint detection. However, due to the existence of repetitive structures, noisy data, and point density variations, these methods hardly achieve a correspondence set completely free of outliers. Therefore, the outlier-robust optimization back-end is highly desirable.

Ideally, if the correspondences are known, the registration problem can have a closed-form solution [Hor87; AHB87], despite the non-convexity of  $\text{SO}(3)$ . Besides, Olsson *et al.* [OKO08] proposed a globally optimal registration method based on branch and bound (BnB) for the outlier-free problem. Their method enables point-to-point, point-to-line, and point-to-plane registration. In practical scenarios, several well-established paradigms have been extensively studied to suppress the outlier correspondences and achieve robust registration, such as *M-estimation* [ZPK16; YSC20; Li+20; SMG23], *outlier removal* [ART10; BC17; Cai+19; Li22;

Yan+22a; Li+23b], *consistency voting* [Gle+14; Yan+19; SSO20; Yan+22b; Yan+23], and *consensus maximization* [FB81; QY20; BM21; Yan+21; Che+22d; Che+23b; Zha+23b]. In addition to the aforementioned geometric-based paradigms, recent studies have also utilized deep learning techniques for correspondence-based point cloud registration, such as 3DRegNet [Pai+20], DGR [CDK20], PointDSC [Bai+21], DHVR [Lee+21], DetarNet [CYT22], and VBReg [Jia+23a].

A representative M-estimation method, called fast global registration (FGR), was proposed by [ZPK16]. This method formulates the registration problem using the Geman-McClure objective function and then combines graduated non-convexity (GNC) to solve the registration problem. While this approach is highly efficient, it can easily generate incorrect solutions at a high outlier rate. Combining the ideas of outlier removal and M-estimation, Yang *et al.* [YSC20] proposed a certifiable and deterministic approach, i.e., truncated least squares estimation and semidefinite relaxation (TEASER). TEASER leverages the translation invariant measurements (TIMs) to decouple the 6-DOF transformation search problem into a 3-DOF rotation search sub-problem followed by a 3-DOF translation search sub-problem. Meanwhile, TEASER allows outlier pruning by maximum clique method [Bus+19], which, however requires quadratic memory space (i.e.,  $\mathcal{O}(N^2)$ ). The pioneering work of outlier removal is [ART10], which proposed an inlier selection method based on the game-theoretic framework for point cloud registration. This method is stochastic and has no global optimality guarantees. Parra and Chin [BC17] first proposed a guaranteed outlier removal (GORE) method, which leverages geometrical bounds to prune outliers and guarantees that eliminated correspondences are not the inlier. Another outlier removal method based on GORE was proposed by [Li22], which is initialization-free and only costs polynomial running time to prune the outlier correspondences. However, it sacrifices global optimality for efficiency. The same authors proposed a quadratic-time GORE [Li+23b], which preserves the global optimality while significantly improving the efficiency. Yan *et al.* [Yan+22a] presented an outlier removal strategy based on the reliability of the correspondence graph, called graph reliability outlier removal (GROR). The pioneering work in the consistency voting paradigm was introduced by Buch *et al.* [Gle+14], which proposed a voting-based method that integrates local and global constraints to score correspondences. Sahloul *et al.* [SSO20] presented a two-stage voting strategy aimed at ranking correspondences based on local and global geometric consistencies. In a more recent study, Yang *et al.* [Yan+23] introduced a mutual voting method that involves calculating the node clustering coefficients of the compatibility graph constructed from the correspondence set. Subsequently, both nodes and edges in the graph are mutually scored, leading to the ranking of correspondences based on the voting scores.

With regard to the consensus maximization paradigm, the most popular registration method is based on the heuristic RANSAC [FB81]. During each iteration, RANSAC employs a minimal solver to calculate the 3-DOF rotation and the 3-DOF translation separately. Nonetheless, RANSAC-based methods demonstrate efficient performance primarily when dealing with low outlier rates. Moreover, they are non-deterministic and generate a correct solution only with a certain probability due to the essence of random sampling [Le+19]. Recently, several RANSAC-based variants [CMK03; LMC12; LHA20; LHA21; BM21; Sun21;

[BNM21] were proposed by introducing novel sampling strategies or local optimization methods. For instance, Graph-cut RANSAC [BM21] (GC-RANSAC) introduced the graph-cut algorithm to improve the local optimization performance. In [QY20], authors introduced a new sampling strategy guided by compatibility to prioritize samples consisting of geometrically compatible correspondences. Yang *et al.* [Yan+21] proposed a consensus maximization method for robust registration, called sample consensus by sampling compatibility triangles in graphs (SAC-COT). They introduced a new correspondence sampling method based on the compatibility triangle representation. More recently, Chen *et al.* [Che+22d] introduced a new metric based on second-order spatial compatibility to assess correspondences and proposed a two-stage sampling strategy to maximize the inlier set accordingly. Besides, Zhang *et al.* [Zha+23b] introduced a point cloud registration method based on maximal cliques. This method first constructs a graph from the correspondence set and searches maximal cliques in the graph. Then, it generates and evaluates the hypotheses according to the maximal cliques and finally selects the best one. Despite previous efforts, these methods still face challenges in terms of time efficiency and accuracy, especially when dealing with high outlier rates.

Against this background, several deterministic global optimization methods have been proposed, most of which are based on the globally optimal BnB framework [BSP12; BC17; Cai+19; Che+22c]. BnB [Sch11; Flo13] is a powerful optimization algorithm framework for solving non-convex and NP-hard problems. The fundamental concept underlying BnB involves the iterative alternation between solution domain segmentation (branch) and sub-branch bounds computation (bound) until the globally optimal solution is obtained. Among the existing BnB-based methods, Bazin *et al.* [BSP12] proposed a globally optimal consensus maximization method based on BnB for the 3-DOF rotation search. Except for the proposed guaranteed outlier removal method, GORE [BC17] converts the 6-DOF registration problem to a 3DOF rotational registration problem and then utilizes BnB to maximize the inlier set. Then, Cai *et al.* [Cai+19] presented an extension of GORE for the 4-DOF terrestrial LiDAR registration. This method jointly searches for the globally optimal solution in the high-dimensional domain by BnB. To speed up the BnB algorithm, Chen *et al.* [Che+22c] proposed both a new transformation decomposition strategy for 6-DOF registration and an efficient two-stage search method based on BnB. Their decomposition strategy is different from current TIMs-based methods in that it decomposes 6-DOF into (2+1)-DOF and (1+2)-DOF by special geometric constraints. Although some transformation decomposition strategies have been developed, the main limitation of the BnB algorithm is the high time cost. Theoretically, the computational complexity of BnB optimization is exponential to the dimension of the solution domain in the worst-case. Hence, there is still potential for improving the algorithm efficiency by considering the dimensionality of the registration problem.

### 2.2.2 Correspondence-free Registration

In contrast to correspondence-based methods, correspondence-free methods utilize raw point data as input instead of features, eliminating the need for explicit correspondence between points. Expectation-maximization (EM)-type methods are considered as classic approaches



for solving the simultaneous pose and correspondence registration (SPCR) problem. These methods aim to iteratively find the maximum likelihood or maximum a posteriori (MAP) estimates of parameters. Within EM-type methods, ICP [BM92] is the most representative and popular. Starting from an initial estimation, it solves the association of correspondence and the estimation of transformation alternately until convergence. ICP also has many variants [RL01; GP02; Fit03; CSK05; SHT09; Mai+11; Pav+18; Rus19; ZYD21; Li+22], which enhance the robustness and efficiency. For example, Trimmed ICP [CSK05] introduced a trimmed least squares cost function. Sparse ICP [BTP13] introduced a sparse  $L_p$  norm to build the loss function. Robust ICP [ZYD21] adapted the Welsch function as the robust kernel. RSICP [Li+22] presented a symmetric point-to-plane distance metric to construct the robust loss function. However, ICP and its variants commonly tend to converge to the local minima during the iterative optimization process and are highly dependent on the initial transformation estimation.

In recent years, researchers have developed several methods of representing the point cloud with Gaussian mixture models (GMMs) to improve robustness [TK04; JV05; MS10; JV10]. These methods transform the registration problem into the probability distribution alignment problem, which seeks to maximize the similarity of two distributions. Although these methods converge quickly to the optimal value when they have reasonable initial estimations, they also do not provide any performance guarantees. Besides, these methods are highly dependent on the initial estimations, and if the initial start point is not properly set, these algorithms tend to converge to the local minimum. Furthermore, there is another kind of heuristic method [AMC08] for the SPCR problem, which is similar to RANSAC, i.e., the 4-points congruent sets approach (4PCS). Rather than sampling a minimal of three points, 4PCS-based registration methods sample four approximately co-planar points as the base set and then find the corresponding four points in the target point cloud. They assume that the cross-diagonal ratios are constant under the affine transformation in any planar quadrangle. Many variants [TWS14; MAM14; Ge17; Xu+19] have also been developed based on 4PCS, with enhancements that include using sparse keypoints as candidate points to reduce the number of input points, increasing robustness, and reducing computational complexity, etc. However, these local methods are not guaranteed to provide optimal solutions.

In addition to these local methods, there has been a great deal of research on solving the SPCR problem with global methods, such as particle swarm optimization [KN10], filtering-based [SDT09; Li+17], and simulated annealing [LLH00]. Methods of this kind have an increased probability of reaching the global optimum regardless of the initialization conditions. However, global optimality cannot be guaranteed. Another direction to globally address the SPCR problem is estimating the globally optimal solution based on the BnB algorithm, such as [LH07; Yan+16; Bus+16; CP16; Str+17; Liu+18b; Cam+18; Liu+18a; Cam+19; Wan+21a]. A pioneer work is [LH07], in which authors proposed a rotational registration method that utilizes BnB to minimize the Lipschitz cost function. Go-ICP [Yan+16] is the first practical globally optimal approach for the 6-DOF SPCR problem that employs the nested BnB search structure to minimize the objective function based on the  $L_2$  residual. Parra *et al.* [Bus+16] introduced a tight bound function based on stereographic projections

to achieve a fast 3-DOF rotation search by BnB. They also extended the proposed method to perform globally optimal 6-DOF registration by utilizing the nested BnB framework. Campbell *et al.* [CP16] proposed a more efficient and accurate BnB-based approach to minimize the GMM-based objective function. However, both these methods jointly search for the globally optimal solution over the 6-dimensional parameter space, leading to relatively high computational costs. One common direction to improve algorithm efficiency is reducing the dimension of the solution domain by decoupling the transformation. For instance, Straub *et al.* [Str+17] proposed a decoupling method based on surface normal distributions, which decomposes the 6-DOF registration problem into the separate 3-DOF rotation and 3-DOF translation sub-problems. Liu *et al.* [Liu+18b] introduced the rotation invariant features (RIFs) to enable sequential estimations of the 3-DOF translation and the 3-DOF rotation instead of the joint 6-DOF transformation search. On the other hand, several BnB-based methods were also proposed for solving the 2D-3D SPCR problem. For instance, Campbell *et al.* [Cam+18] modeled the 2D-3D SPCR problem as a consensus maximization problem and proposed tight bound functions to jointly search for the 6-DOF solution. Liu *et al.* [Liu+18a] also used the consensus maximization scheme to robustly solve the 2D-3D rotation search problem and derived several different bound functions for BnB. Wang *et al.* [Wan+21a] leveraged RIFs to decouple the joint 6-DOF searching into two separate 3-DOF searching processes and proposed efficient bound functions. In general, these global approaches still suffer from inefficiencies, especially on data with large sizes and high outlier rates.

Recently, some investigations also tried to achieve end-to-end registration based on deep learning [Aok+19; WS19a; WS19b; YL20; Yua+20; Fu+21; YL22]. For instance, PointNetLK [Aok+19] proposed an end-to-end registration model based on the Lucas and Kanade (LK) algorithm. DCP [WS19a] utilized feature similarity to establish soft correspondences and subsequently employed weighted SVD to calculate the rigid transformation. PR-Net [WS19b] utilized keypoint detection to solve the partial-to-partial problem. RPM-Net [YL20] employed the Sinkhorn normalization and annealing to achieve soft assignments of correspondences. DeepGMR [Yua+20] leveraged a probabilistic registration paradigm based on GMM. RegTR [YL22] used attention mechanisms to achieve significant feature representation. However, learning-based methods require additional training procedures, and their generalization capabilities are not always reliable.



# 3

## Globally Optimal Robust Radar Calibration

*Radar is among the most popular sensors in modern Intelligent Transportation Systems (ITSs), enabling weather-robust perception. The orientation and position of the traffic radar are crucial for perception fusion in ITSs. However, due to the unknown target association, sparseness, and noisiness of radar measurements, robust and accurate extrinsic calibration is challenging. This chapter proposes a targetless traffic radar calibration method based on GPS to overcome the inconvenience during ITS operation. Furthermore, in the back-end of extrinsic calibration, this chapter proposes a globally optimal registration method, which is highly robust to noise and outliers in radar measurements. Extensive experiments demonstrate that the proposed method is not only globally optimal but also more accurate and robust.*

### 3.1 Background

Modern Intelligent Transportation Systems can significantly improve transportation efficiency and reduce the occurrence of traffic accidents by providing vehicles with extra detailed information about the present traffic flow in the form of a digital twin [Zha+11; Krä+19]. In ITSs, many heterogeneous sensors, such as monocular cameras, event cameras, LiDARs, radars, etc., should be equipped to acquire information on different modalities. These sensors, which usually have different capabilities and fields of view, need to be integrated into the ITS. Extrinsic calibration, a key step in multi-sensor integration, can find the spatial relationship between the sensor coordinate systems and build a consistent environment model [Dub+14; Du+21; Bel+22; Jia+23b]. The focus of this chapter is the sensor extrinsic calibration in ITSs.

Traffic radar is known as the most popular sensor for wide applications in ITSs, such as object detection, tracking, and localization [CA15; Wan+16; Arn+20; Che+20]. Because the millimeter wave can penetrate fog, smoke, and dust easily, the millimeter wave radar is robust, especially in extreme weather conditions [Sko80; LMD21]. Moreover, it can detect targets at longer distances than other sensors, such as LiDARs. However, the measurements of traffic radar have the following properties:

- Because of the low spatial resolution in the vertical direction, the field of view of the traffic radar is considered to be planar.
- Traffic radar detections are sparse, as shown in Fig. 3.1. Because the traffic radar outputs measurements solely in the form of detected objects (traffic participants).
- Traffic radar measurements are usually contaminated by noise and outliers because the reflected waves come from all moving and non-moving targets.

Therefore, due to the low precision and high outlier rate of radar measurements, accurate and robust extrinsic calibration of traffic radar is crucial but challenging. On the other hand, because traffic radar is used to detect and track traffic participants, mainly to confirm which lane they are in, the accuracy of traffic radar calibration is required at the meter level.

Traditional extrinsic calibration methods are generally classified as target-based [PMP17; Oh+18; PMP19] method and targetless [KD20; Pen+20; Yua+21] method. Among the target-based methods, researchers leverage external dedicated targets, such as the corner reflector, to build reference features for detection and association. However, target-based calibration methods are infeasible for the real practical scenario, such as during ITS operation. Because the installation of dedicated calibration targets on the highway is impractical and dangerous. In addition, it is not always practical to calibrate the sensor with a prepared target for different real applications. Thus, to overcome the inconvenience during ITS operation, this thesis proposes a targetless extrinsic calibration method based on GPS for traffic radar. The global and accurate positioning information from high-precision GPS can improve the accuracy of the calibration. The extrinsic calibration problem is then transformed into estimating the rigid transformation between the measurement sets of radar and GPS. The destination measurement set is obtained from the GPS mounted on the moving vehicle, and the source measurement set is obtained from the traffic radar mounted on the ITS infrastructure. By combining with the high-precision GPS, the proposed traffic radar calibration method can promote the environmental perception fusion among each measurement point in the ITS so as to build an accurate global coordinate system for the digital twin.

Please note that, for both target-based and targetless calibration methods, the complex target association problem must be addressed. In most studies, the extrinsic calibration is transformed into a correspondence-based registration problem [AHB87; BC17; YSC20]. For instance, the correspondences are derived from the dedicated calibration target in target-based calibration approaches. In targetless approaches, researchers use environmental features sensed by both sensors to find the correspondences, such as geometric descriptors. Nevertheless, traffic radars can not provide any descriptive features such as edges and corners because of the measurement mechanism and the limited resolution. The association of radar measurements with vehicles in images or geometrical features from LiDAR point clouds is intractable. This association problem must be rethought from a different perspective, such as using the registration method to find correspondences between measurements from different sensors. Therefore, the above-mentioned alignment of the measurement sets from radar and GPS can be considered as a simultaneous pose and correspondence registration (SPCR) problem [BM92; BS03; MS10; JV10]. Different from the correspondence-based registration



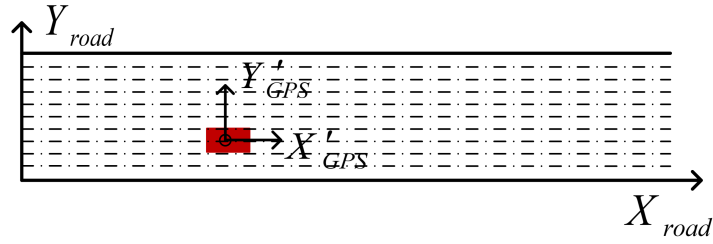
**Figure 3.1:** This is the image after the traffic radar measurements being projected onto the camera picture. The measurements (blue points) are sparse and contain biases and misdetections caused by noise.

problem, the correspondences in this problem are unknown, such that the correspondences and transformation need to be addressed simultaneously. Moreover, the SPCR problem is usually non-convex due to the quadratic orthogonality constraints of rotation [Yan+16]. Existing methods, such as ICP [BM92], CPD [MS10], and GMMReg [JV10], have been used to solve the SPCR problem. However, they are prone to delivering erroneous results (local optimum) during the iterative optimization process and are highly dependent on the initial estimation of transformation, i.e., initialization.

Recently, globally optimal methods have been proposed according to the branch and bound (BnB) paradigm [Cla99; Sch11; TN20]. The prominent features of the BnB-based method are that 1) it can find the globally optimal solutions with theoretical guarantee, 2) it is independent of initialization. In this study, a robust globally optimal method is proposed to solve the SPCR problem for targetless calibration, called Gaussian Mixture Robust Branch and Bound (GMRBnB). In detail, the measurement sets are first represented by the Gaussian mixture model (GMM) [Rey09; JV10] to counteract the influence of noise and outliers. Next, the robust objective function is constructed using the closed-form solution of the statistical difference metric between two Gaussian mixtures. Then, a BnB-based optimization method with newly proposed relaxation bounds is presented to find the globally optimal transformation. Notably, due to the inherent high computational cost drawback of BnB-based methods [Cam+18], the proposed method is only used for offline calibration, which is admissible in real applications. This study puts more attention on the accuracy of the calibration, i.e., how to find the globally optimal solution.

The main contributions of this chapter can be summarized as follows:

- This chapter proposes a targetless calibration method for traffic radar to tackle the inconvenience during ITS operation. The combination of high-precision GPS on the moving vehicle and traffic radar on the ITS infrastructure enables accurate extrinsic



**Figure 3.2:** The top view of GPS coordinate system.

calibration. Besides, there are no restrictions on the relative position of sensors, i.e., the relative angle and translation are arbitrary.

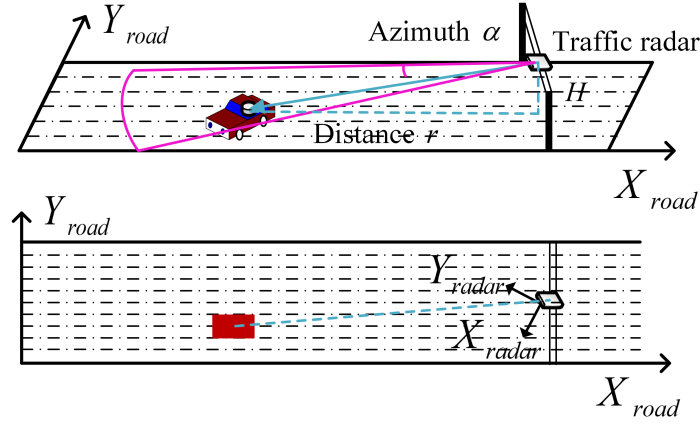
- This chapter proposes a BnB-based globally optimal registration method to align the measurements of traffic radar and GPS in a simultaneous pose and correspondence registration manner. The GMRBnB algorithm searches for the globally optimal solution in the domain of 2D rotation and translation. Therefore, the proposed method can avoid failure by getting trapped in local optima in the case of large relative angles.
- Extensive experiments demonstrate that the proposed method achieves accurate and robust performance on both synthetic and real-world data in contrast to the existing methods.

The rest of this chapter is organized as follows: Section 3.2 illustrates the problem formulation of the traffic radar extrinsic calibration. Section 3.3 demonstrates the principle and details of the proposed method. Section 3.4 presents the experiment results with synthetic data and real-world data from the Providentia++ project. Finally, Section 3.5 gives the summary.

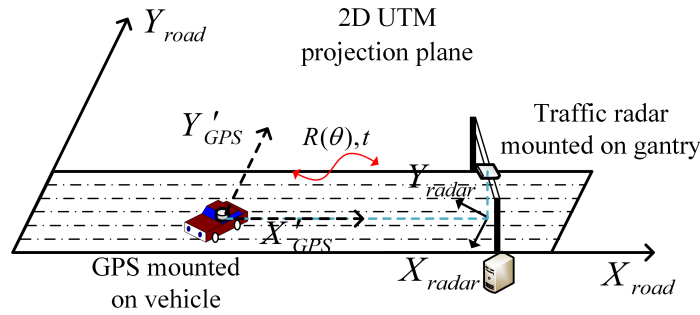
## 3.2 Problem Formulation

### 3.2.1 ITS Coordinate System

The measurement of GPS is the position in the WGS-84 coordinate system, i.e.,  $(X_{GPS}, Y_{GPS}, Z_{GPS})$ . It also is the trajectory of the moving vehicle with timestamps. In this study, the coordinate system of GPS measurements is converted to the same planar universal transverse mercator (UTM) coordinate system as the ITS coordinate system  $(X_{road}, Y_{road})$ , i.e.,  $(X'_{GPS}, Y'_{GPS})$ , as shown in Fig. 3.2. Further, radar detection of targets is all based on the same principle: low-frequency electromagnetic pulses are emitted from the radar antenna and reflected back to the sensor from conductive targets in the environment [Sko80]. The bearing, distance, and radial velocity of the target can be obtained by measuring the time of flight and phase of the returned pulse. The roadside traffic radar performs projection on a horizontal plane with depth and azimuth of a detected target, and the projected point is denoted as  $m_r(\alpha, r)$  in the 2D polar coordinate system, where  $\alpha$  and  $r$  are azimuth and depth of the target respectively, as shown in Fig. 3.3. The height of the traffic radar  $H$  can be easily measured



**Figure 3.3:** The traffic radar coordinate system and its top view.



**Figure 3.4:** The ITS coordinate system related to the extrinsic calibration.

by a laser rangefinder so that the measurements of traffic radar can be projected onto the same UTM plane with the coordinate system  $(X_{\text{radar}}, Y_{\text{radar}})$ , as shown in Fig. 3.3. The overall ITS coordinate system is shown in Fig. 3.4. Regarding the traffic radar measurements, as explained above, only the pre-processed data of detected targets can be acquired. Hence, under the geometric representation, the measurements in the measurement set are represented as discrete points in the coordinate system  $(X_{\text{radar}}, Y_{\text{radar}})$ . Therefore, the proposed targetless extrinsic calibration method is transformed into aligning the two measurement sets of radar and GPS in the 2D UTM projection plane such that the relative pose of traffic radar can be obtained.

### 3.2.2 Measurement Set Registration Problem

Define the source and destination measurement sets  $\mathcal{X} = \{X_i\}_{i=1}^m$  and  $\mathcal{Y} = \{Y_j\}_{j=1}^n$ , where  $X_i, Y_j \in \mathbb{R}^2$  represent the coordinates of measurements in radar and GPS frames, respectively. The problem is estimating the correspondence and 3-DOF rigid transformation (with rotation angle  $\theta \in [-\pi, \pi]$  and translation  $\mathbf{t} \in \mathbb{R}^2$ ) between the two measurement sets, so that the transformed set  $\mathcal{X}$  is as close as possible to the set  $\mathcal{Y}$ . The  $L_2$  residual between the transformed point  $X_i$  and point  $Y_j$  is defined as

$$e_{X_i, Y_j}(\theta, \mathbf{t}) = \|\mathbf{R}(\theta)X_i + \mathbf{t} - Y_j\|_2 \quad (3.1)$$

where  $\mathbf{R}(\theta)$  is the rotation matrix, i.e.,

$$\mathbf{R}(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (3.2)$$

Further, the residual between the transformed point  $\mathbf{X}_i$  and set  $\mathcal{Y}$  is defined as

$$e_{\mathbf{X}_i}(\theta, \mathbf{t}) = \min_{\mathbf{Y}_j \in \mathcal{Y}} e_{\mathbf{X}_i, \mathbf{Y}_j}(\theta, \mathbf{t}) \quad (3.3)$$

which is to find the correspondence between point  $\mathbf{X}_i$  and set  $\mathcal{Y}$ , i.e., find the closest point of  $\mathbf{X}_i$  from the set  $\mathcal{Y}$ . Then the objective function to align the two measurement sets is given by

$$E(\theta, \mathbf{t}) = \sum_{i=1}^m e_{\mathbf{X}_i}(\theta, \mathbf{t})^2 \quad (3.4)$$

The SPCR problem is trying to simultaneously find the optimal solution and correspondences that minimizes the objective function in Eq. (3.4). If the real correspondences in Eq. (3.3) are known, the rigid transformation in Eq. (3.4) can be calculated in the closed-form by the optimizing algorithms such as the least squares optimization method based on singular value decomposition (SVD). However, it is difficult to determine correspondences perfectly in practical applications. The solution of traditional local methods is prone to converge to the local minimum for the SPCR problem. On the other hand, the results are susceptible to noise and outliers in measurement sets. In the following section, the GMM is utilized to construct the negative objective function for enhancing the robustness and propose a BnB-based method to find the globally optimal solution for this SPCR problem.

### 3.3 Method

#### 3.3.1 Gaussian Mixture Model and Similarity Measures

The representation of measurement sets with GMM is extensively employed to solve the registration problem [CP15; CP16; MS10; JV10; Hir21; LWC21]. The main idea of this representation is mapping the discrete measurement sets to the continuous domain by probability density function, which is defined as a weighted sum of Gaussian density functions, i.e.,  $p(\mathbf{x}|\mathcal{G}) = \sum_{i=1}^k \omega_i \mathcal{N}(\mathbf{x}|\mu_i, \sigma_i^2)$ . It is the probability density of observing a point  $\mathbf{x}$  given a GMM  $\mathcal{G} = \{\omega_i, \mu_i, \sigma_i^2\}_{i=1}^k$ , with mixture weights  $\omega_i$ , means  $\mu_i$ , variances  $\sigma_i^2$  and the number of Gaussian components  $k$ , respectively. In summary, the benefits of GMM representation include: 1) it uses a continuous probability density field for the representation of measurements and does not require partitioning and discretization of the space, 2) the covariance matrix in the GMM can better deal with the noise and outliers in the measurements, and 3) the correspondence update or closest point search is not performed during the optimization process.

Assuming no prior information, a simple construction method for GMM from the given

measurement set is as follows: 1) all Gaussian components have the same weights and covariance matrixes, 2) the mean of each component is the spatial coordinate of the corresponding measurement point, and 3) the number of Gaussian components and the measurement points are identical. Then the GMMs  $\mathcal{G}_x = \{\omega_i^x, \mathbf{X}_i, \sigma_{i\mathcal{X}}^2\}_{i=1}^m$  and  $\mathcal{G}_y = \{\omega_j^y, \mathbf{Y}_j, \sigma_{j\mathcal{Y}}^2\}_{j=1}^n$  are generated from the destination and source measurement sets  $\mathcal{X}$  and  $\mathcal{Y}$  respectively, as shown in Fig. 3.5.

In this study, the  $L_2$  distance is employed as the statistical difference metric between two GMMs since the formulation can be expressed in closed-form [JV10; CP16]. The rigid transformation function  $T(\mathcal{G}, \theta, \mathbf{t})$  is defined to denote the rotation  $\theta \in [-\pi, \pi]$  and translation  $\mathbf{t} \in \mathbb{R}^2$  for the GMM. The  $L_2$  distance between two GMMs is defined as

$$D(\theta, \mathbf{t}) = \int_{\mathbb{R}^2} \left( p(\mathbf{x}|T(\mathcal{G}_x, \theta, \mathbf{t})) - p(\mathbf{x}|\mathcal{G}_y) \right)^2 d\mathbf{x} \quad (3.5)$$

Eq. (3.5) is then expanded as follows:

$$D(\theta, \mathbf{t}) = \int_{\mathbb{R}^2} \left[ \left( p(\mathbf{x}|T(\mathcal{G}_x, \theta, \mathbf{t})) \right)^2 + \left( p(\mathbf{x}|\mathcal{G}_y) \right)^2 - 2p(\mathbf{x}|T(\mathcal{G}_x, \theta, \mathbf{t}))p(\mathbf{x}|\mathcal{G}_y) \right] d\mathbf{x} \quad (3.6)$$

where the first term is invariant for any rigid transformation, the second term is independent of the rigid transformation, and the third term is relevant to the rigid transformation. However, the third term has a closed-form, derived by the following formula:

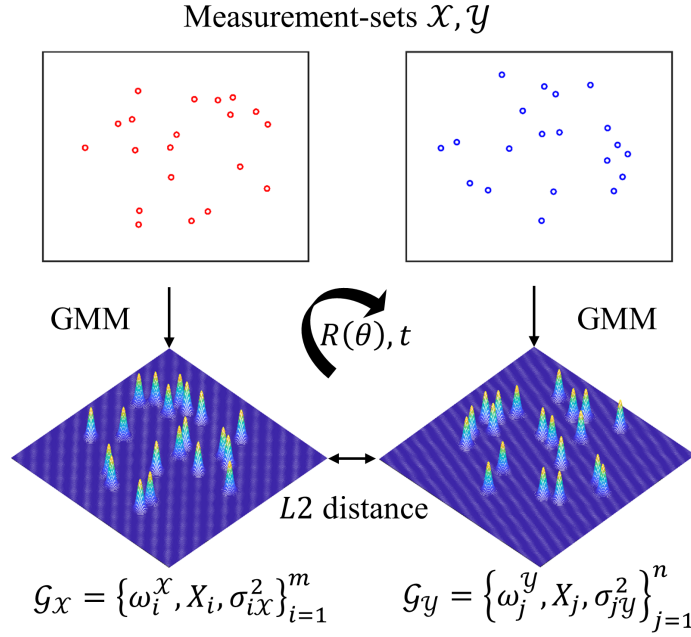
$$\begin{aligned} & \int_{\mathbb{R}^2} \mathcal{N}(\mathbf{x}|\mu_1, \sigma_1^2) \mathcal{N}(\mathbf{x}|\mu_2, \sigma_2^2) d\mathbf{x} \\ &= \mathcal{N}(\mathbf{0}|\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2) \end{aligned} \quad (3.7)$$

Thus the negative GMM robust objective function over the  $L_2$  distance is given by

$$\begin{aligned} G(\theta, \mathbf{t}) &= - \int_{\mathbb{R}^2} p(\mathbf{x}|T(\mathcal{G}_x, \theta, \mathbf{t})) p(\mathbf{x}|\mathcal{G}_y) d\mathbf{x} \\ &= - \sum_{i=1}^m \sum_{j=1}^n \omega_i^x \omega_j^y \mathcal{N}(\mathbf{0}|\mathbf{R}(\theta)\mathbf{X}_i + \mathbf{t} - \mathbf{Y}_j, \sigma_{i\mathcal{X}}^2 + \sigma_{j\mathcal{Y}}^2) \\ &= - \sum_{i=1}^m \sum_{j=1}^n \frac{\omega_i^x \omega_j^y}{N} \exp \left[ - \frac{[e_{\mathbf{X}_i, \mathbf{Y}_j}(\theta, \mathbf{t})]^2}{2[\sigma_{i\mathcal{X}}^2 + \sigma_{j\mathcal{Y}}^2]} \right] \end{aligned} \quad (3.8)$$

where  $e_{\mathbf{X}_i, \mathbf{Y}_j}(\theta, \mathbf{t})$  is the point-to-point  $L_2$  residual in Eq. (3.1) and  $N$  is the normalisation factor. A negative GMM objective function is built instead of the traditional objective function in Eq. (3.4), which displays statistical robustness to counteract the noise and outliers. Further, the most significant difference is that the GMM objective function does not need to find the closest corresponding point, such as Eq. (3.3). The function is the sum of the sums, not the sum of the minima so that the problem to be solved is simplified. In the next section, the relaxation bounds for the minimum of this objective function are derived.





**Figure 3.5:** The representation of measurement sets  $\mathcal{X}$  and  $\mathcal{Y}$  with GMMs  $\mathcal{G}_X$  and  $\mathcal{G}_Y$ .

### 3.3.2 BnB and Relaxation Bounds

BnB is a global optimization framework for solving non-convex and NP-hard problems. It searches the entire solution domain to seek a globally optimal solution with theoretical guarantees. The BnB-based algorithm operates according to two principles: 1) branch, it recursively divides the search domain into smaller sub-boxes, and 2) bound, it evaluates the lower and upper bounds on the optimal solution in each constrained sub-box. Then, the algorithm uses these bounds to prune the search domain and delete the sub-box that can be proved not to contain the optimal solution. The algorithm converges when the lower and upper bounds on the optimal solution are tight enough, i.e., a predetermined threshold is achieved. The algorithm depends on efficient estimation of the lower and upper bounds of branches. Breadth-first search (BFS) and depth-first search (DFS) have been employed to traverse the tree of sub-boxes [MSS08]. The DFS strategy is recommended when no good heuristic is available for producing an initial estimation.

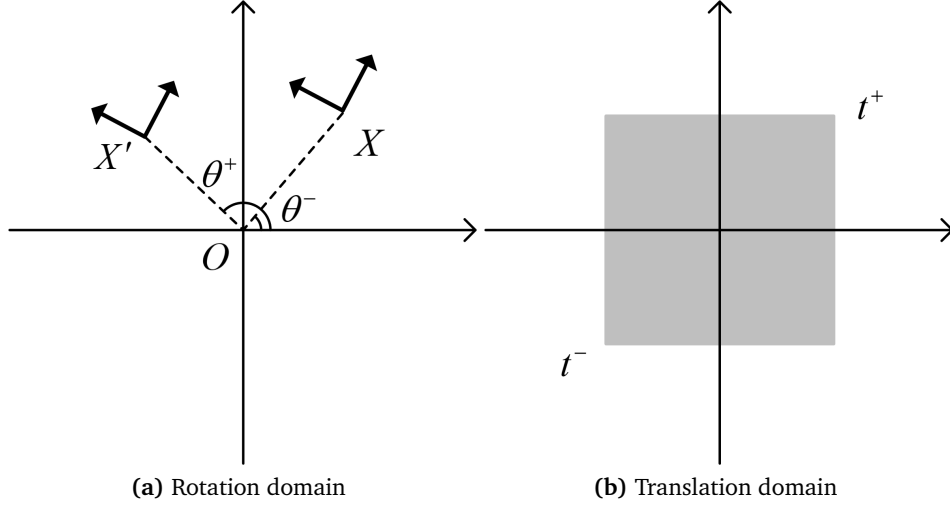
Define  $\mathbb{D}$  is the search domain and  $\mathbb{B} \subset \mathbb{D}$  is the segmented sub-box. Assume that functions  $f_L(\mathbb{B})$  and  $f_U(\mathbb{B})$  satisfy

$$f_L(\mathbb{B}) \leq \min_{\forall(\theta, t) \in \mathbb{B}} G(\theta, t) \leq f_U(\mathbb{B}) \quad (3.9)$$

where  $f_L(\mathbb{B})$  is the lower bound function and  $f_U(\mathbb{B})$  is the upper bound function. It should be noted that the upper bound function can generally be chosen as the value obtained by substituting any value in the search domain or sub-boxes into the objective function. Besides, Eq. (3.10) is the property of bounds functions that can guarantee the convergence of the BnB-based algorithm, which is given by

$$\lim_{\sigma(\mathbb{B}) \rightarrow 0} (f_L(\mathbb{B}) - f_U(\mathbb{B})) = 0 \quad (3.10)$$





**Figure 3.6:** Search domain parametrization for BnB. The rotation domain is the range of rotation angle, i.e.,  $[\theta^-, \theta^+]$ . The translation is assumed to be a 2D rectangle  $[t^-, t^+]^2$ . The splitting policy is to divide from the center of each dimension of the domain, so there will be  $2^3 = 8$  sub-boxes after one splitting.

where  $\sigma(\mathbb{B})$  is the diameter of sub-box  $\mathbb{B}$ . Eq. (3.9) and Eq. (3.10) described above are necessary conditions to prove the bounds functions.

According to the registration problem, the parameterization of the search domain is  $\mathbb{D} = \{(\theta, \mathbf{t}) \in [-\pi, \pi] \times \mathbb{R}^2 \mid t^- \leq \mathbf{t} \leq t^+\}$ . The rotation domain and translation domain are shown in Fig. 3.6. The next step is then to derive the lower and upper bound of the  $L_2$ -based negative GMM objective function  $G(\theta, \mathbf{t})$  within the domain  $\mathbb{D}$ , which is the core of the BnB-based algorithm. In the objective function  $G(\theta, \mathbf{t})$ , all terms except for the point-to-point  $L_2$  residual are easy to calculate. Thus the focus is on bounding the point-to-point residual  $e_{X_i, Y_j}(\theta, \mathbf{t})$ , as shown in the following proposition.

**Proposition 3.1** (Relaxation bounds of the objective function  $G(\theta, \mathbf{t})$ ). *For the sub-box  $\mathbb{B} = \{(\theta, \mathbf{t}) \mid \theta^- \leq \theta \leq \theta^+, t^- \leq \mathbf{t} \leq t^+\}$  centred at  $(\theta_c, \mathbf{t}_c)$ , the relaxation upper bound  $R_U(\mathbb{B})$  and lower bound  $R_L(\mathbb{B})$  of the negative GMM robust objective function  $G(\theta, \mathbf{t})$  for  $X_i, Y_j$  are*

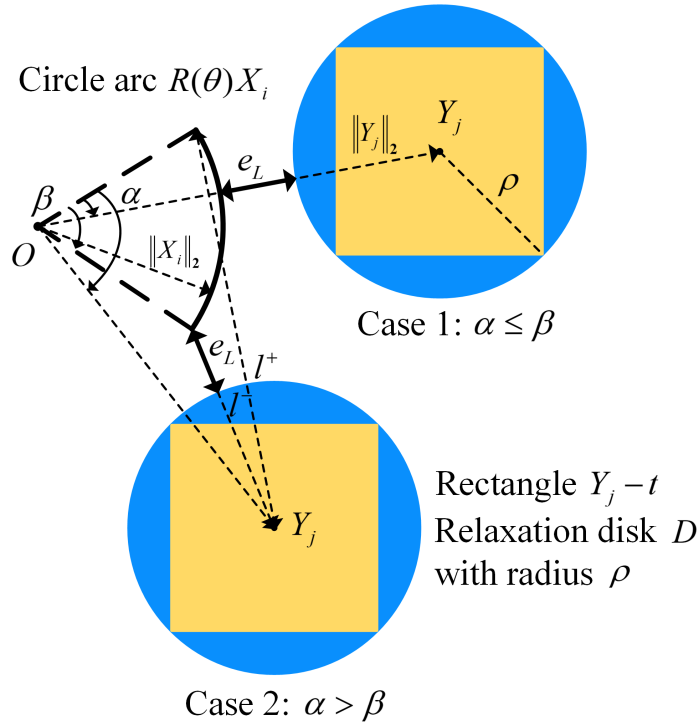
$$R_U(\mathbb{B}) = - \sum_{i=1}^m \sum_{j=1}^n \frac{\omega_i \omega_j}{N} \exp \left[ - \frac{[e_{X_i, Y_j}(\theta_c, \mathbf{t}_c)]^2}{2[\sigma_i^2 + \sigma_j^2]} \right] \quad (3.11)$$

$$R_L(\mathbb{B}) = - \sum_{i=1}^m \sum_{j=1}^n \frac{\omega_i \omega_j}{N} \exp \left[ - \frac{[e_{L, X_i, Y_j}(\theta, \mathbf{t})]^2}{2[\sigma_i^2 + \sigma_j^2]} \right] \quad (3.12)$$

where  $e_{L, X_i, Y_j}(\theta, \mathbf{t})$  is the lower bound of the point-to-point  $L_2$  residual  $e_{X_i, Y_j}(\theta, \mathbf{t})$ , and is given by

$$e_{L, X_i, Y_j}(\theta, \mathbf{t}) = \begin{cases} \max \{ \|\mathbf{Y}_j\|_2 - \|\mathbf{X}_i\|_2 - \rho, 0 \}, & \alpha \leq \beta \\ \max \{ \min \{ l^-, l^+ \} - \rho, 0 \}, & \alpha > \beta \end{cases} \quad (3.13)$$

where angles  $\alpha, \beta$  are shown in Fig. 3.7,  $\rho$  is the radius of the translation relaxation disk, and  $l^- = \|\mathbf{Y}_j - \mathbf{R}(\theta^-)\mathbf{X}_i\|_2$ ,  $l^+ = \|\mathbf{Y}_j - \mathbf{R}(\theta^+)\mathbf{X}_i\|_2$ .



**Figure 3.7:** The schematic of the relaxation lower bound function  $R_L(\mathbb{B})$  with two cases, and  $e_L$  is the abbreviation of the lower bound of the point-to-point residual  $e_{L, X_i, Y_j}(\theta, \mathbf{t})$ .

*Proof.* To prove the effectiveness and validity of the relaxation lower bound  $R_L(\mathbb{B})$  and upper bound  $R_U(\mathbb{B})$  for the objective function  $G(\theta, \mathbf{t})$ , there are three issues that need to be addressed.

- Observe that  $\forall(\theta, \mathbf{t}) \in \mathbb{B}$ ,

$$\min e_{X_i, Y_j}(\theta, \mathbf{t}) = \min \|\mathbf{R}(\theta)\mathbf{X}_i - (\mathbf{Y}_j - \mathbf{t})\|_2 \quad (3.14)$$

This minimum can be denoted as the minimum distance between circular arc  $\{\mathbf{R}(\theta)\mathbf{X}_i | \theta \in [\theta^-, \theta^+]\}$  and rectangle  $\{\mathbf{x} \in \mathbb{R}^2 | \mathbf{Y}_j - \mathbf{t}^+ \leq \mathbf{x} \leq \mathbf{Y}_j - \mathbf{t}^-\}$  in any branch sub-box  $\mathbb{B} = \{(\theta, \mathbf{t}) | \theta^- \leq \theta \leq \theta^+, \mathbf{t}^- \leq \mathbf{t} \leq \mathbf{t}^+\}$ , and the radius of the arc is  $\|\mathbf{X}_i\|_2$ . Then the rectangle domain can be relaxed to a disk shaped domain as shown in Fig. 3.7 and can be expressed by the following equation.

$$D = \{\mathbf{x} \in \mathbb{R}^2 | \|\mathbf{x} - \mathbf{Y}_j\|_2^2 - \rho^2 \leq 0\} \quad (3.15)$$

where  $\rho$  is the radius of the relaxation disk, and  $\rho = \|\mathbf{t}^+ - \mathbf{t}^-\|_2 / 2$ . Obviously, the minimum distance between the circular arc and the rectangle is greater than or equal to the minimum distance between the circular arc and the relaxation disk, while both are greater than or equal to zero, that is,

$$\min e_{X_i, Y_j}(\theta, \mathbf{t}) \geq \min \|\mathbf{R}(\theta)\mathbf{X}_i - D\|_2 \geq 0 \quad (3.16)$$

Then the minimum distance between the circular arc and the relaxation disk is equal

to the minimum distance between the circular arc and the center of the disk domain  $Y_j$  minus the radius of the relaxation disk  $\rho$ , which is

$$\min \|\mathbf{R}(\theta)\mathbf{X}_i - D\|_2 = \min \|\mathbf{R}(\theta)\mathbf{X}_i - Y_j\|_2 - \rho \quad (3.17)$$

The minimum distance between the circular arc and the center of the disk domain is an easy problem, that is, the minimum distance between the arc and the point. According to the different position relationships between the arc and the point, it is divided into two cases for consideration, as shown in Fig. 3.7. When the point lies within the rotation sector corresponding to the circular arc (Case 1), the minimum distance between the arc and the point is the difference between the two distances from the origin, which is  $|\|Y_j\|_2 - \|X_i\|_2|$ . When the point lies outside the rotation sector corresponding to the circular arc (Case 2), the minimum distance between the arc and the point is the minimum distance between the endpoints of the arc  $\mathbf{R}(\theta^-)\mathbf{X}_i$ ,  $\mathbf{R}(\theta^+)\mathbf{X}_i$  and the point, which is the minimum of  $\{l^-, l^+\}$ . Details are shown below,

$$\begin{aligned} & \min \|\mathbf{R}(\theta)\mathbf{X}_i - Y_j\|_2 - \rho \\ &= \begin{cases} \max\{\|Y_j\|_2 - \|X_i\|_2 - \rho, 0\}, & \alpha \leq \beta \\ \max\{\min\{l^-, l^+\} - \rho, 0\}, & \alpha > \beta \end{cases} \end{aligned} \quad (3.18)$$

As for angles  $\alpha$  and  $\beta$ , since the endpoints of the circular arc  $\mathbf{R}(\theta^-)\mathbf{X}_i$ ,  $\mathbf{R}(\theta^+)\mathbf{X}_i$  and the center of the disk domain  $Y_j$  are known, it is easy to compare these two angles. Define  $e_{L, X_i, Y_j}(\theta, \mathbf{t}) = \min \|\mathbf{R}(\theta)\mathbf{X}_i - Y_j\|_2 - \rho$ , then,

$$\min_{\forall(\theta, \mathbf{t}) \in \mathbb{B}} e_{X_i, Y_j}(\theta, \mathbf{t}) \geq e_{L, X_i, Y_j}(\theta, \mathbf{t}) \quad (3.19)$$

By substitution into Eq. (3.8),

$$\min_{\forall(\theta, \mathbf{t}) \in \mathbb{B}} G(\theta, \mathbf{t}) \geq R_L(\mathbb{B}) \quad (3.20)$$

Therefore,  $R_L(\mathbb{B})$  is the lower bound of the objective function  $G(\theta, \mathbf{t})$ .

- The point-to-point  $L_2$  residual at a specific point within any sub-box of the search domain is larger than the minimal residual within the search domain, that is

$$e_{X_i, Y_j}(\theta_c, \mathbf{t}_c) \geq \min_{\forall(\theta, \mathbf{t}) \in \mathbb{B}} e_{X_i, Y_j}(\theta, \mathbf{t}) \quad (3.21)$$

Then substituting into Eq. (3.8),

$$R_U(\mathbb{B}) \geq \min_{\forall(\theta, \mathbf{t}) \in \mathbb{B}} G(\theta, \mathbf{t}) \quad (3.22)$$

Therefore,  $R_U(\mathbb{B})$  is the upper bound of the objective function  $G(\theta, \mathbf{t})$ .

- When the sub-box  $\mathbb{B} = \{(\theta, \mathbf{t}) | \theta^- \leq \theta \leq \theta^+, \mathbf{t}^- \leq \mathbf{t} \leq \mathbf{t}^+\}$  collapses to a single point

---

**Algorithm 1:** GMRBnB: An algorithm for globally optimal spatial registration with Gaussian Mixture Model representation

---

**Input:** Gaussian Mixture Models  $\mathcal{G}_X, \mathcal{G}_Y$  with means  $X_i$  and  $Y_j$  respectively, weights  $\omega$ , and variances  $\sigma^2$ ; Solution domain  $\mathbb{D}$ ; Convergence threshold  $\epsilon$ .

**Output:**  $\epsilon$ -optimal solution  $(\theta, \mathbf{t})^*$ .

- 1 Let  $\xi$  be the list of sub-boxes, initialize  $\xi = \{\mathbb{D}\}$ , and normalize means  $X_i$  and  $Y_j$ ;
  - 2 **while** *The tolerance between global upper and lower bound is larger than the threshold  $\epsilon$ ,  $U - L > \epsilon$*  **do**
  - 3     Select a sub-box  $\mathbb{B}$  with the minimum of lower bound, i.e.,  $\mathbb{B} = \arg \min R_L(\mathbb{B}_k)$ ,  $\mathbb{B}_k \in \xi$ , and split it into eight sub-boxes  $S(\mathbb{B}) = \{\mathbb{B}_1, \dots, \mathbb{B}_8\}$ ;
  - 4     Delete  $\mathbb{B}$  from  $\xi$ , and add  $\{\mathbb{B}_1, \dots, \mathbb{B}_8\}$  into  $\xi$ ;
  - 5     Update  $L = \min R_L(\mathbb{B}_k), \mathbb{B}_k \in \xi$ ;
  - 6     Update  $U = \min \{U, R_U(\delta(\mathbb{B}_k))\}$  with  $\mathbb{B}_k \in \xi$ . If  $U = R_U(\delta(\mathbb{B}_k))$ , set  $(\theta, \mathbf{t})^* = \delta(\mathbb{B}_k)$ ;
  - 7     Delete  $\mathbb{B}_k$  from  $\xi$  with  $R_L(\mathbb{B}_k) > U$ ;
  - 8 **end**
- 

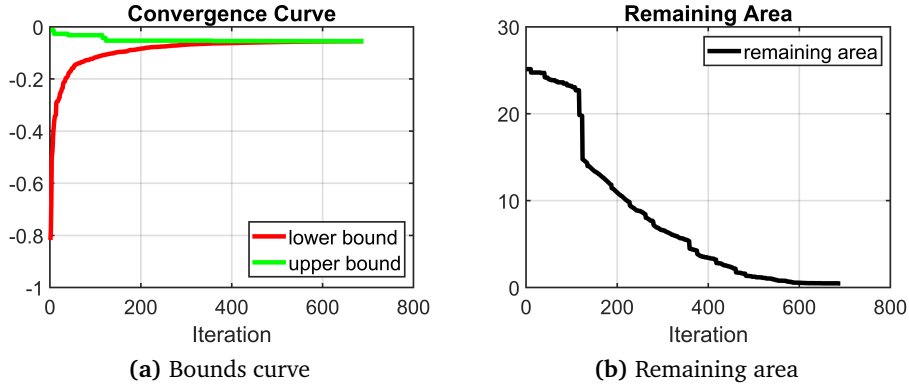
$(\theta_0, \mathbf{t}_0)$ , then,  $e_{X_i, Y_j}(\theta_0, \mathbf{t}_0) = e_{L, X_i, Y_j}(\theta_0, \mathbf{t}_0)$  and  $R_L(\theta_0, \mathbf{t}_0) = R_U(\theta_0, \mathbf{t}_0)$ , i.e.,

$$\lim_{\sigma(\mathbb{B}) \rightarrow 0} (R_L(\mathbb{B}) - R_U(\mathbb{B})) = 0 \quad (3.23)$$

The gap between the upper and lower bound is equal to zero, and the convergence is proved. □

### 3.3.3 GMRBnB Algorithm

According to the relaxation bounds of the negative GMM robust objective function  $G(\theta, \mathbf{t})$  in the last section, the GMRBnB algorithm is outlined in Algorithm 1. To simplify the calculation and speed up the algorithm, GMMs are assumed to have the same covariance matrix and weights. In practice, the coordinates of the measurement set can be normalized so that the translation domain can be  $[-1, 1]^2$ , which can not only achieve a smaller box but also ensure that the domain covers every feasible translation, as shown in Line 1. Define function  $\delta(\mathbb{B}) \in \mathbb{B}$  returns the center point of box  $\mathbb{B}$ , and function  $S(\mathbb{B})$  divides the box  $\mathbb{B}$  in half in each dimension of the domain. The exploration policy in Algorithm 1 is DFS. In detail, the search priority is inverse to the value of the lower bound, as shown in Line 3, and the next branch to be divided is the branch with the minimal lower bound. Line 4 is intended to delete the box that has been split. Line 5 is intended to update  $L$  to the minimum of the current lower bound of all branches. Line 6 is intended to update  $U$  to the minimum of the upper bound of all branches in all iterations. The pruning policy is in Line 7, and the algorithm will prune the branch whose lower bound is larger than the global upper bound. After the algorithm converges, the  $\epsilon$ -optimal solution is obtained without correspondences, i.e., only



**Figure 3.8:** Convergence curves of the proposed GMRBnB algorithm.

the transformation parameters  $(\theta, \mathbf{t})$ . This is because the GMM-based objective function avoids solving the explicit target association problem, i.e., finding the closest corresponding point. In addition, the algorithm outputs the best-so-far solution in each iteration, as shown in Line 6. Therefore, it can still provide best-guess transformation parameters when dealing with problems with limited running time, even if the upper and lower bounds are not fitted tightly enough.

## 3.4 Experiments

To demonstrate the performance and global optimality of the proposed method, GMRBnB is compared with existing registration algorithms, including ICP [BM92], CPD [MS10], and GMMReg [JV10], using challenging synthetic and real data in this section. These algorithms are conducted in MATLAB 2019B, and all experiments are implemented on a laptop with an i7-9750H CPU and 16GB RAM.

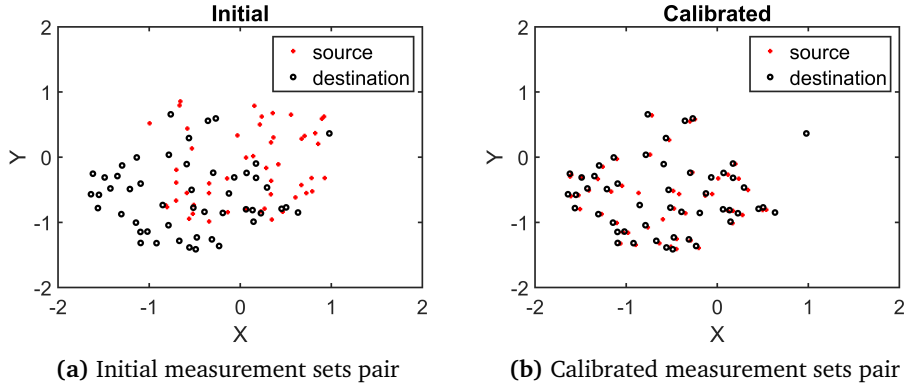
### 3.4.1 Experimental Setup and Convergence

According to the property of the GMM objective function, the convergence range is influenced by the standard deviation  $\sigma$  of the Gaussian components, i.e., the width of the peak corresponding to the Gaussian distribution. Therefore, after some trying and comparing, the standard deviation  $\sigma$  of the Gaussian components is set to 0.1 in all experiments. What's more, there exists a normalization factor  $N$  in the objective function  $G(\theta, \mathbf{t})$ . Thus, the convergence threshold for GMRBnB is set as  $\epsilon = 0.01$ . In addition, to evaluate the accuracy and robustness, the translation error and rotation error are defined as

$$e_{\mathbf{t}} = \|\mathbf{t}_{\text{gt}} - \mathbf{t}^*\| \quad (3.24a)$$

$$e_{\theta} = \|\theta_{\text{gt}} - \theta^*\| \quad (3.24b)$$

where  $\mathbf{t}_{\text{gt}}$  and  $\theta_{\text{gt}}$  are motion ground truth,  $\mathbf{t}^*$  and  $\theta^*$  are optimal solutions.



**Figure 3.9:** An example of measurement sets pair before and after calibration.

For a simple demonstration of the convergence for the proposed lower and upper bounds, a pair of synthetic measurement sets ( $m = n = 20$ ) is set as input, and the convergence curves are obtained in Fig. 3.8. It is obvious that the gap between the lower and upper bounds is converging to zero, and after several hundred iterations, the proposed method can converge to the optimal solution. Moreover, the remaining area is rapidly decreasing.

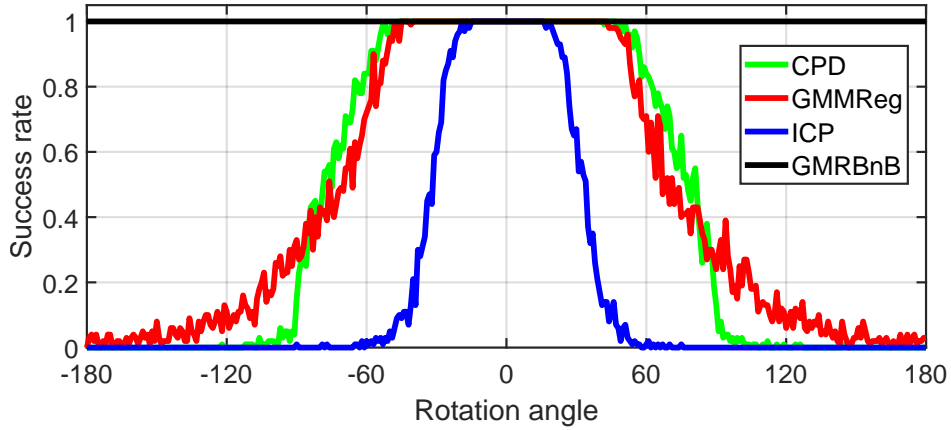
### 3.4.2 Control Experiments on Synthetic Data

This section illustrates the accuracy, robustness, and global optimality of the proposed algorithm relative to the other three algorithms through three control experiments. Synthetic data with different experimental conditions containing rotation angle, outlier rate, and noise level is employed. In the experiments, the false negatives and false positives in the correspondences are denoted as outliers uniformly.

#### Data generation

Initially, the first measurement set is generated by creating  $m$  random points that are distributed in square  $[-1, 1]^2$ . Then, the random rotation in  $[-\pi, \pi]$  and random translation in  $[-1, 1]^2$  are applied to this measurement set to obtain the corresponding transformed measurement set. They are the source measurement set and the destination measurement set, respectively. An example of a measurement sets pair before and after calibration is shown in Fig. 3.9. The simulation of the outliers is performed by randomly replacing some points in the transformed measurement set. The simulation of the noise is achieved by perturbing the measurement set with the noise from the uniform distribution  $U[-\delta_{\text{noise}}, \delta_{\text{noise}}]$ .

Meanwhile, the average of the translation error  $e_t$  and rotation error  $e_\theta$  for  $T$  trials represent the accuracy. The median runtime of  $T$  trials under each set of experimental conditions is also recorded in order to compare the efficiency and computational expense. Further, to demonstrate the global optimality of the proposed method, the success rate is defined as  $T^+/T$ , where  $T$  is the total number of trials under the same experimental conditions and  $T^+$  is the number of successful cases satisfying  $e_t < 0.1\text{m}$  and  $e_\theta < 5^\circ$ . As for the experimental



**Figure 3.10:** Success rate of CPD, GMMReg, ICP, and GMRBnB algorithms without translation.

conditions, the outlier rate is  $\lambda = m_{\text{outlier}}/m$  where  $m_{\text{outlier}}$  is the number of outliers and  $m$  is the total number of points in the measurement set. Further, the value of  $\delta_{\text{noise}}$  is considered as the noise level.

#### Deterministic global optimality

In this section, the deterministic global optimality of the proposed GMRBnB algorithm is compared with ICP, CPD, and GMMReg. In terms of the details of data generation, the source measurement set is rotated in the range of  $[-180^\circ, 180^\circ]$  at  $1^\circ$  increments without translation to generate the destination measurement set. Each measurement sets pair has  $m = n = 50$  points, and the outlier rate and noise level are  $\lambda = 0$  and  $\delta_{\text{noise}} = 0$  separately. The measurement sets pair is randomly generated 100 times for each rotation angle, and then the related registration experiments with different algorithms are performed. A case is considered successful when the rotation error satisfies  $e_\theta < 5^\circ$ . The same experiments for ICP, CPD, and GMMReg are performed, and the success rate versus rotation angle is shown in Fig. 3.10.

As seen from Fig. 3.10, the proposed GMRBnB algorithm maintains a 100% success rate over the entire range of rotation angles, which represents the deterministic global optimality. However, the other algorithms only maintain a 100% success rate over a small range of rotation angles. The ICP algorithm has a 100% success rate only when rotation angle is in the range  $[-15^\circ, 15^\circ]$ , the CPD algorithm has a 100% success rate only in the range  $[-48^\circ, 48^\circ]$ , and the GMMReg algorithm has a 100% success rate only in the range  $[-45^\circ, 45^\circ]$ . When the range is exceeded, the success rate of these algorithms decreases rapidly. Moreover, when the rotation angle is close to about  $\pm 60^\circ$ , the success rate of the ICP algorithm is close to zero, that for the CPD algorithm is about  $\pm 105^\circ$ , and that for the GMMReg algorithm is about  $\pm 150^\circ$ . The overall success rates of both the GMMReg and CPD algorithms are higher than that of the ICP algorithm, which indicates that the ICP algorithm is easier to converge to a local optimum. In contrast, the GMRBnB algorithm has deterministic global optimality, and its convergence range is the whole rotation domain, which is much better than that of the other algorithms.

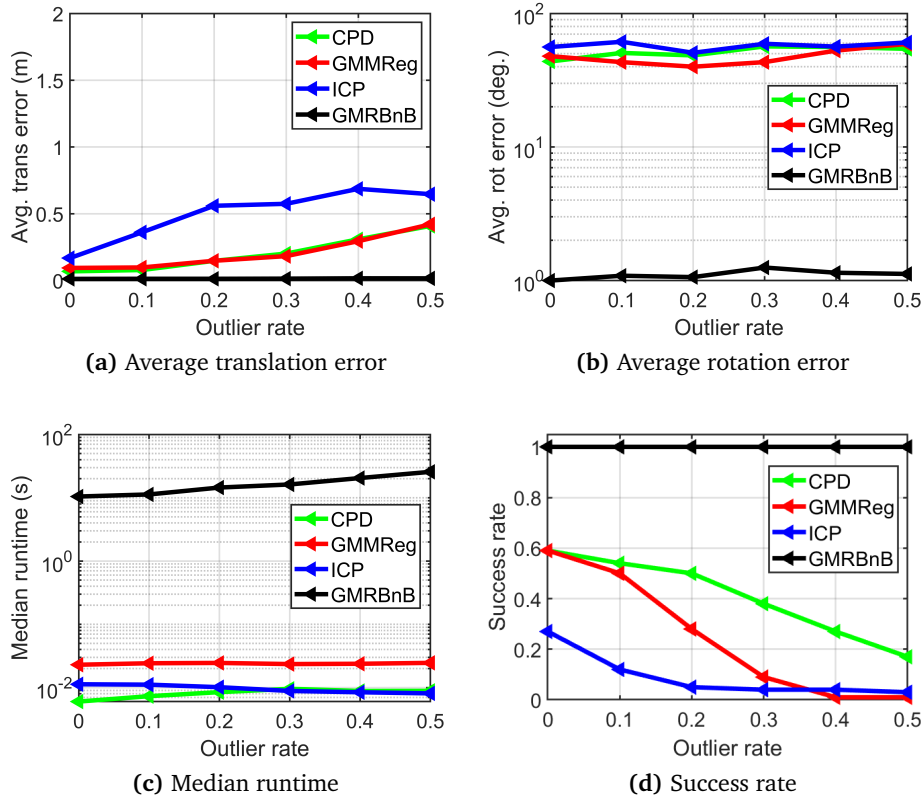


Figure 3.11: Control experiments on synthetic data with different outlier rates.

### Robustness to noise and outliers

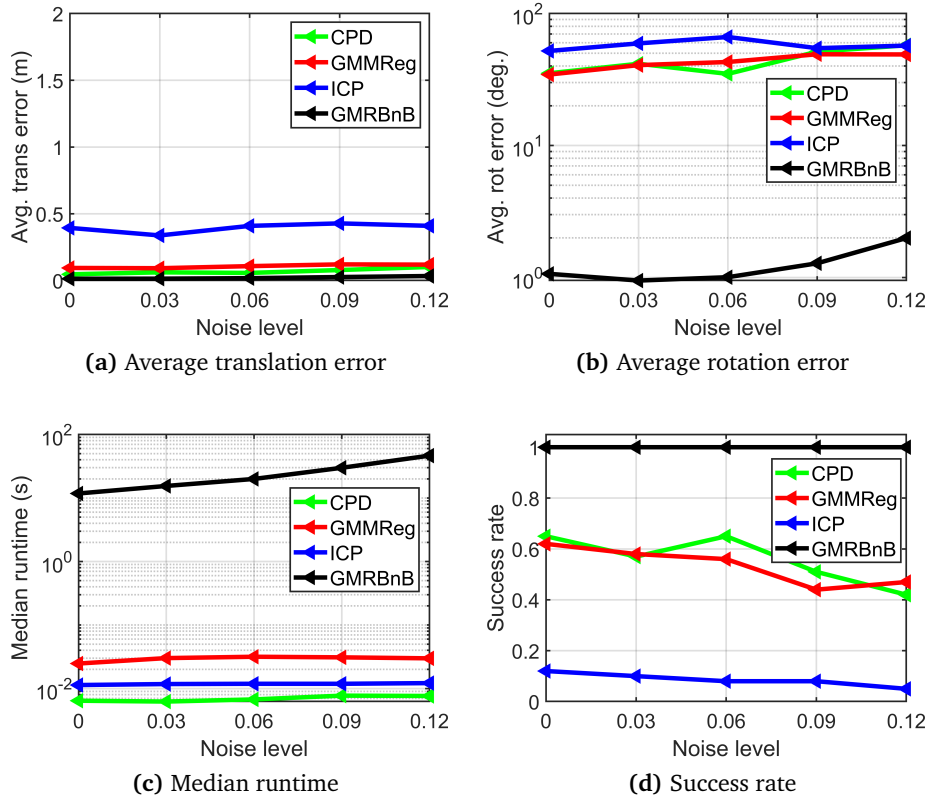
In this section, the robustness to noise and outliers of each algorithm is compared by challenging synthetic data experiments. The experimental conditions include different outlier rates and noise levels. Firstly, the robustness of the GMRBnB algorithm is tested in various outlier rates  $\lambda = \{0, \dots, 0.5\}$  and then compared with other algorithms. The pair of measurements set with  $m = n = 50$  points is generated in each experiment, and the noise level is  $\delta_{\text{noise}} = 0.1$ . Under each experimental setting, the experiment is repeated  $T = 100$  times to verify the generality and global optimality. Further, the average rotation error, average translation error, median runtime, and success rate are recorded and plotted in Fig. 3.11.

Secondly, the robustness of the GMRBnB algorithm is tested at different noise levels  $\delta_{\text{noise}} = \{0, \dots, 0.12\}$ . The pair of measurements set with the same  $m = n = 50$  points is generated, and the outlier rate is  $\lambda = 0.1$ . This experiment is also repeated  $T = 100$  times in each experimental setting to observe the global optimality. Besides, the average error, median runtime, and success rate are plotted in Fig. 3.12.

From the results of all control experiments in this section, the summary is given as follows:

- Firstly, as the outlier rate and the noise level increase, the proposed method always maintains a 100% success rate, which confirms the global optimality of the algorithm. Moreover, the robustness of the algorithm against outliers and noise is also demonstrated by the average error.





**Figure 3.12:** Control experiments on synthetic data with different noise levels.

- Secondly, the average rotation errors of the ICP, CPD, and GMMReg algorithms are much larger than that of the GMRBnB algorithm under all experimental conditions. According to the experiments in the previous section, it is clear that the rotation angle has a large effect on the success rate of these algorithms, while translation has a relatively small effect. Therefore, the average rotation errors of these algorithms become large in the arbitrarily repeated experiments. Besides, the average translation errors of the GMRBnB are all smaller than those of the ICP, CPD, and GMMReg algorithms. This also indicates that the proposed method can obtain better accuracy than other methods.
- Finally, the proposed method is higher than other algorithms in terms of computational cost. BnB-based methods search the whole solution domain for the globally optimal solution, which is essential for safety-critical applications but is admissible for offline calibration. Besides, the required time of the proposed method increases with the growth of the outlier rate or noise level.

### 3.4.3 Calibration Experiments on Real-World Data

In this section, the performance of GMRBnB on real-world data is investigated. These real data are collected by driving the test vehicle on parts of the German highway A9, which is the test field of ITS. The test vehicle is equipped with a high-precision RTK-GPS device, as shown

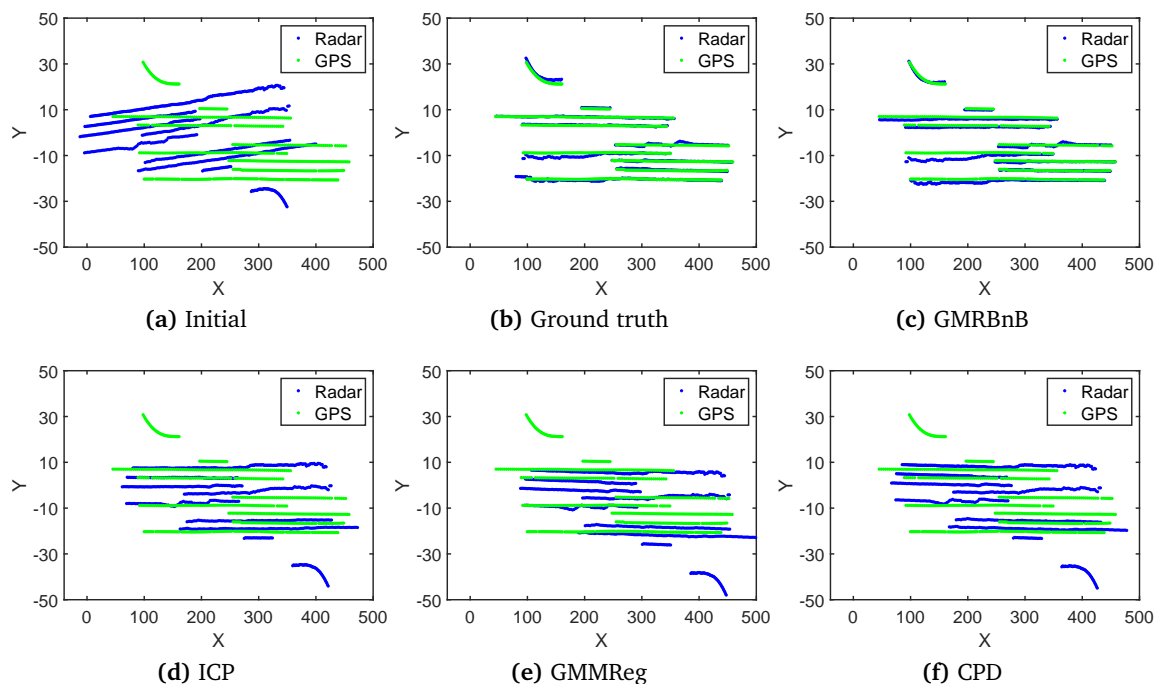


**Figure 3.13:** This is the test car with a high-precision RTK-GPS device (cm-level).

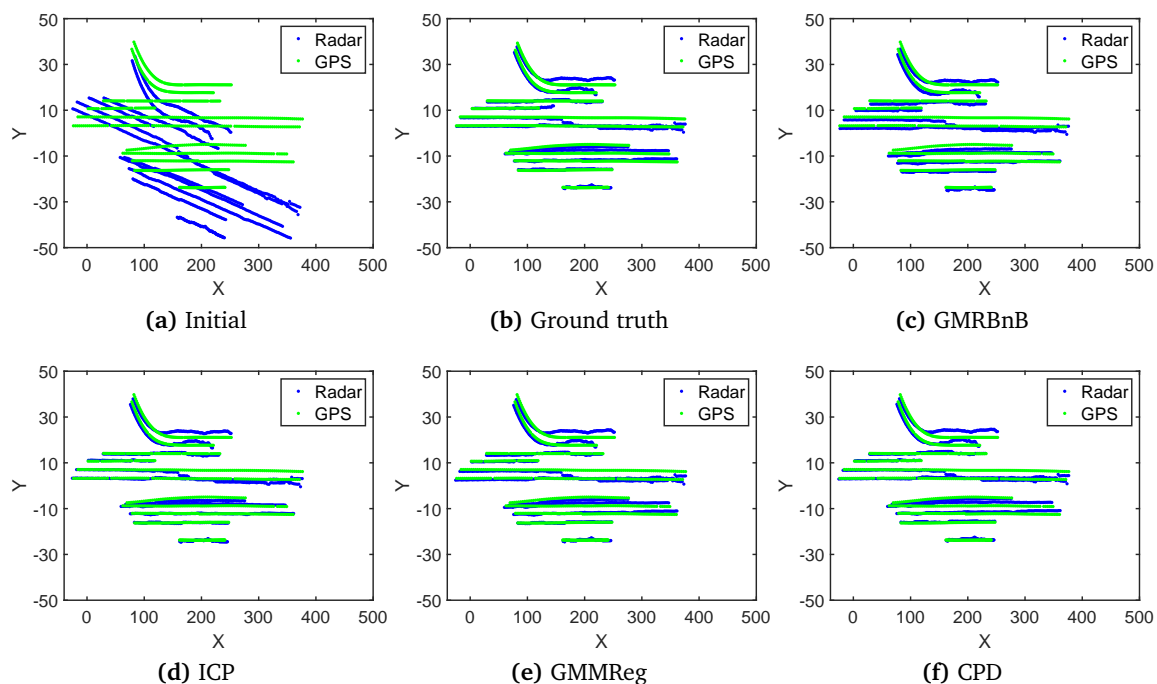
in Fig. 3.13. Multiple sensors are installed on the highway infrastructures, specifically on gantry bridges, overlooking a total of ten traffic lanes. We repeatedly drove on the highway several times, each time passing in a different traffic lane and collecting detection data on the test vehicle from three traffic radars (Radar RA, RB, and RC). These radars are installed in different locations and directions at two measurement points, but they have overlapped fields of view. Notably, the radar measurements contain missed detections, ghost detections, and multiple detections for trucks or buses due to measurement noise. In addition, the corresponding positioning information of the test vehicles is collected from the RTK-GPS device. The radar measurements are represented as the source measurement set with  $m$  measurement points, and the GPS measurements are represented as the destination measurement set with  $n$  measurement points.

### Manual association and ground truth acquisition

In order to obtain the ground truth poses of these radars, the closed-form solution of transformation is calculated using the method that manually establishes the target correspondence. Specifically, the timing clocks of traffic radar and GPS are unified to the ITS public clock under the form of the UNIX timestamp first. Their measurements are then temporal synchronized by a data interpolation method since they have different sampling frequencies, which are 10Hz and 13Hz, respectively. Besides, the sensor detection is transformed into the ITS coordinate system. The measurements of the two sensors are then manually associated based on UNIX timestamps, and the pair of measurement sets with the same size is obtained, i.e.,  $m = n$ . Finally, the transformation matrix is calculated using the least squares optimization method based on SVD. The results are employed as the transformation ground truth of follow-up experiments, as shown in Fig. 3.14(b), Fig. 3.15(b), and Fig. 3.16(b). It is apparent that the ground truth obtained even by the manual association method does not enable a perfect alignment of the two measurement sets. Thus, realizing the traffic radar extrinsic calibration in this practical situation is challenging.



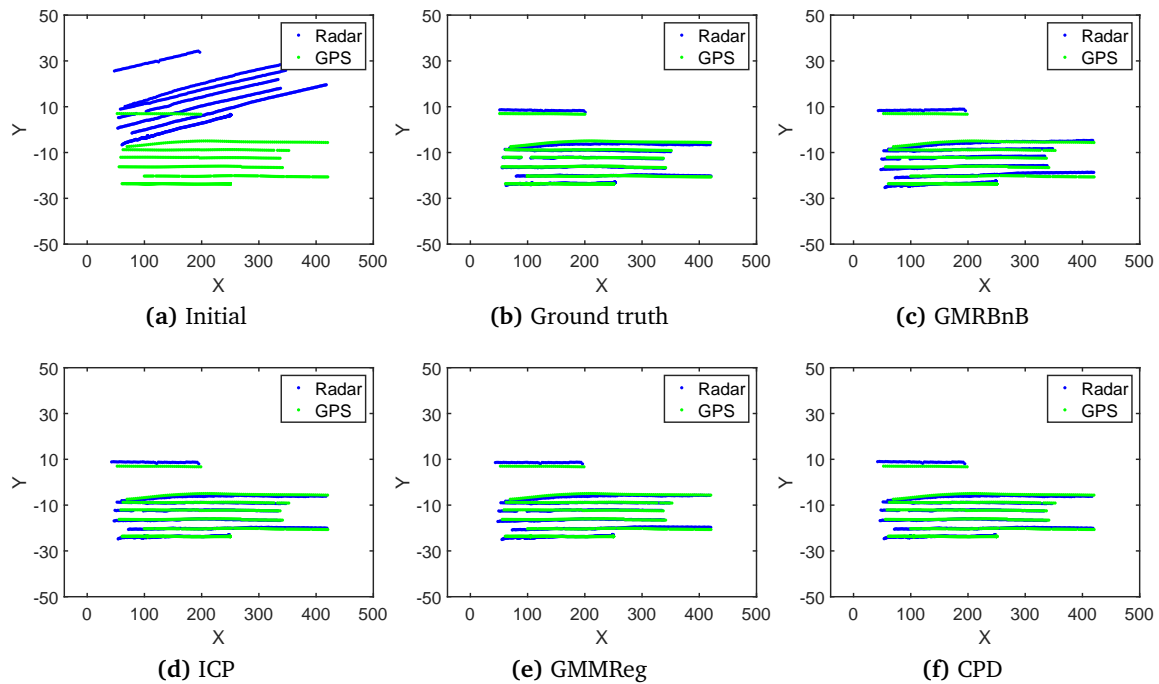
**Figure 3.14:** The initial measurement sets of radar and GPS are  $m = 851$  and  $n = 637$ , respectively. The calibration results of manual association, GMRBnB, ICP, GMMReg, and CPD for Radar RA are compared.



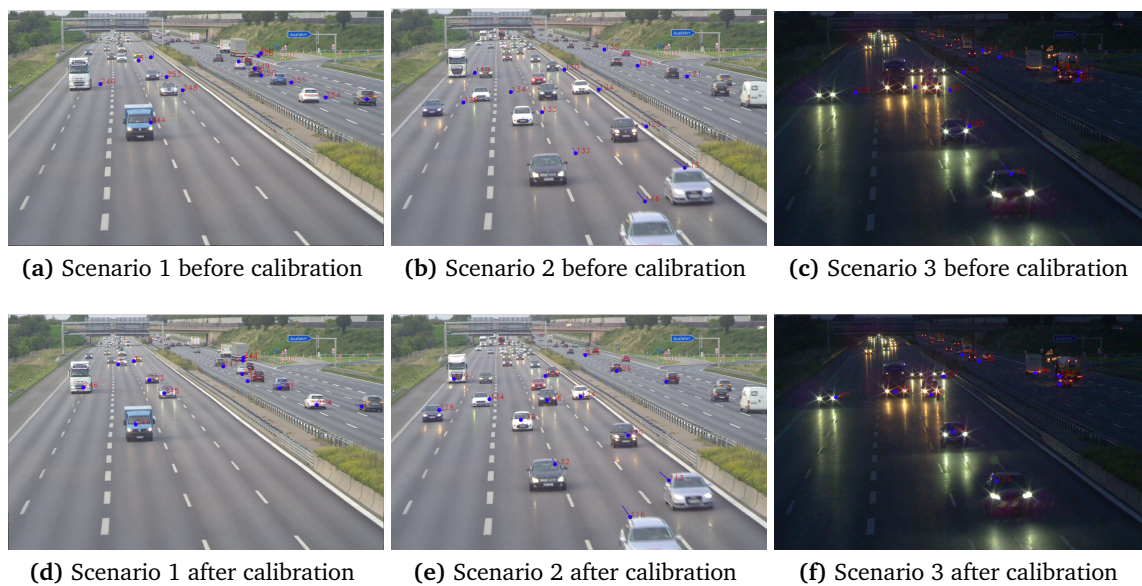
**Figure 3.15:** The initial measurement sets of radar and GPS are  $m = 1199$  and  $n = 896$ , respectively. The calibration results of manual association, GMRBnB, ICP, GMMReg, and CPD for Radar RB are compared.

### Calibration results

In this section, the input data for the experiments are different from the input data for the least squares optimization method in the previous section, which is the real measurement



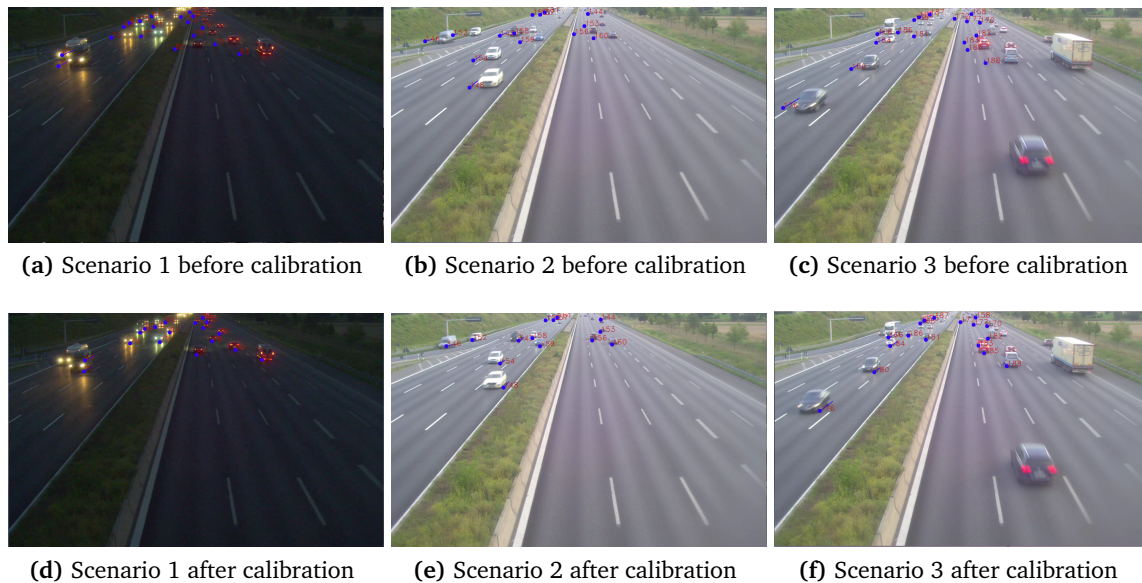
**Figure 3.16:** The initial measurement sets of radar and GPS are  $m = 1122$  and  $n = 881$ , respectively. The calibration results of manual association, GMRBnB, ICP, GMMReg, and CPD for Radar RC are compared.



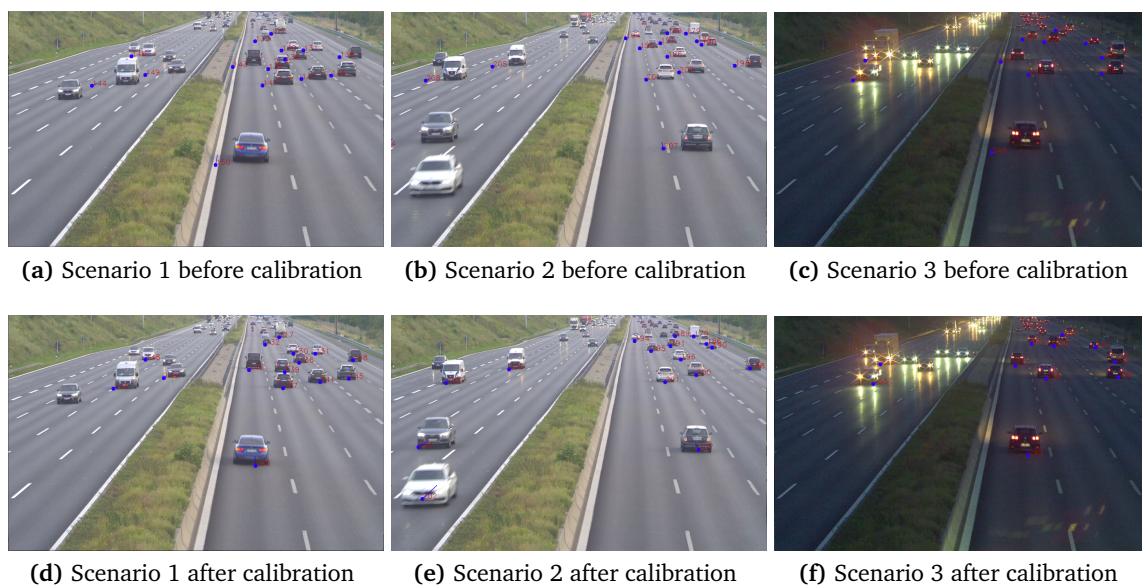
**Figure 3.17:** After projecting the detection of Radar RA (blue points) onto the camera image, the performance of the proposed calibration method is compared before and after its application in three different scenarios.

without manually established correspondence. The size of the measurement sets of the two sensors is different because their sampling rates are different. During the implementation, the input measurement sets are normalized such that the translation domain is set to be  $[-1, 1]^2$ . The proposed targetless calibration method is conducted offline. The accuracy and robustness of the extrinsic calibration is our main concern. Then the calibration results of GMRBnB, ICP, GMMReg, and CPD for three radar are compared with the manually obtained ground truth,





**Figure 3.18:** After projecting the detection of Radar RB (blue points) onto the camera image, the performance of the proposed calibration method is compared before and after its application in three different scenarios.



**Figure 3.19:** After projecting the detection of Radar RC (blue points) onto the camera image, the performance of the proposed calibration method is compared before and after its application in three different scenarios.

as shown in Fig. 3.14, Fig. 3.15, and Fig. 3.16. According to the ground truth, the rotation angles of the data corresponding to the three radars are  $177.65^\circ$ ,  $6.33^\circ$ ,  $3.56^\circ$ , respectively. As can be seen from Fig. 3.14, ICP, GMMReg, and CPD algorithms fail to converge to the optimal solution. However, the proposed method can still obtain the globally optimal solution in the case of a large relative angle. In Fig. 3.15 and Fig. 3.16, all methods, including local methods, converge to the correct result because the relative rotation angle is close to zero degrees. These results demonstrate that the proposed method can find the globally optimal solution with performance guarantees while existing local methods are prone to incorrect solutions

**Table 3.1:** Comparison of rotation error and translation error of all algorithms for three radars

Methods		GMRBnB	ICP [BM92]	GMMReg [JV10]	CPD [MS10]
Radar RA	$e_\theta(^{\circ})$	<b>0.2419</b>	-	-	-
	$e_t(\text{m})$	<b>0.2746</b>	-	-	-
Radar RB	$e_\theta(^{\circ})$	<b>0.0056</b>	12.5305	12.8170	0.0688
	$e_t(\text{m})$	<b>0.9370</b>	2.0060	1.7647	1.6873
Radar RC	$e_\theta(^{\circ})$	<b>6.7084</b>	7.0121	6.9147	7.0350
	$e_t(\text{m})$	8.9213	9.5871	<b>8.5321</b>	10.1841

(local optima) in the case of a large relative angle. The deterministic global optimality of the proposed method is verified in these real-world experiments.

Meanwhile, rotation errors and translation errors of all methods are recorded to compare the accuracy and robustness, as shown in Table 3.1. It is obvious that the rotation error and translation error of GMRBnB are almost the smallest with the data of all three radars. Even for the data of Radar RC, the translation error of GMRBnB is also competitive compared to the smallest value of GMMReg. Compared to other methods, the proposed method has a different formulation, in other words, a different objective function, which is the negative GMM objective function. Theoretically, each objective function has a different surface or shape, leading to slightly different optimal solution locations corresponding to different objective functions [Cam+18]. The solution of the proposed method is theoretically guaranteed to be optimal for the formulation. This is the reason that the translation error of the proposed approach is not minimal for Radar RC. On the other hand, the employed ground truth is manually obtained, and it does not even enable perfect alignment of the two measurement sets due to a large amount of noise and outliers inherent in the traffic radar measurements. Thus, the errors of all methods for Radar RC are relatively large. In summary, the results demonstrate that the proposed method is more robust to outliers and noise than existing local methods.

In addition, three scenarios for each radar from the real recorded data are selected to further demonstrate the performance of the proposed calibration method in the application of camera and traffic radar fusion. After projecting the radar measurements onto the camera image, the performance of the proposed calibration method is compared before and after its application, as shown in Fig. 3.17, Fig. 3.18, and Fig. 3.19. The upper row of images shows the initially erroneous calibration results, where the projected radar measurements do not align with the vehicles in the camera image. After implementing the proposed targetless extrinsic calibration, each detection is overlapped with the corresponding object in the camera image. Notably, a few measurement points are still not perfectly aligned with the vehicles in the image, caused by the noise in measurements. The traffic radar and camera fusion results show that the proposed method achieves precise and reliable extrinsic calibration of traffic radar, and the calibration results are effective for camera and radar fusion.

### 3.5 Summary

In order to tackle the inconvenience during ITS operation, this chapter proposes a targetless extrinsic calibration method for the traffic radar in this chapter. This method does not require any dedicated target and is also safe and easy to implement. On the other hand, a simultaneous pose and correspondence registration method is used to address the difficult target association problem for radar measurements. Further, since existing registration methods are prone to converge to the local optimum and are overly dependent on the initialization, an initialization-free GMRBnB algorithm is proposed to find the globally optimal solution with performance guarantees. The novel relaxation upper and lower bound functions are derived for the GMRBnB algorithm. Then, comparative experiments are conducted on challenging synthetic data to illustrate the theoretical performance of the proposed algorithm. Besides, the performance of the proposed method is evaluated on real-world data and is validated in the application of radar-camera fusion. Extensive experiments demonstrate that the proposed method not only can avoid failures due to getting trapped in local optima, but also is more accurate and robust than existing methods.

Nevertheless, there is still potential space for the improvement of the proposed method in some aspects. Currently, the proposed method is only suitable for offline calibration and is not purely automatic for the ITS. Therefore, achieving automatic calibration and improving efficiency is valuable. Because the sensors are susceptible to environmental disturbances, such as the vibration of the mounting bar due to high winds or displacement due to temperature changes. These perturbations lead to changes in the pose of sensors, hence the calibration parameters need to be automatically updated online. The second point is to further explore the geometrical property in the pose estimation problem and find a new optimization framework, which can effectively accelerate the calibration process.





# 4

## Efficient and Robust LiDAR Registration

*Estimating the rigid transformation between two LiDAR scans through putative 3D correspondences is a typical point cloud registration paradigm. Current 3D feature matching approaches, including learning-based methods, commonly lead to substantial outlier correspondences. Nonetheless, a common drawback of outlier-robust global registration methods is their inherent high computational cost, particularly when dealing with large-scale data and high outlier rates. This chapter introduces a novel pose decoupling strategy based on residual projections to effectively decompose the original registration problem into three sub-problems, thus improving efficiency. Subsequently, a novel BnB-based deterministic search method is proposed to solve these sub-problems within a lower-dimensional domain, resulting in robust and efficient registration. Extensive experiments demonstrate that the proposed method outperforms state-of-the-art methods in terms of efficiency while simultaneously maintaining robustness.*

### 4.1 Background

Rigid point cloud registration is a core and fundamental problem in the field of 3D vision and robotics with a wide range of applications, such as autonomous driving [Zha+22], 3D reconstruction [BL95; Li22], and simultaneous localization and mapping (SLAM) [CVV22; ZS15; Li+23b]. Given the source and target point clouds in different coordinate systems, it aims to estimate the 6 degrees of freedom (DOF) transformation in  $\mathbb{SE}(3)$  to align the two point clouds best. The 6-DOF transformation includes both 3-DOF rotation in  $\mathbb{SO}(3)$  and 3-DOF translation in  $\mathbb{R}^3$ .





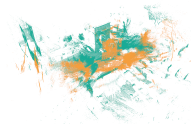
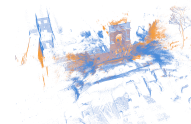
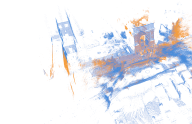






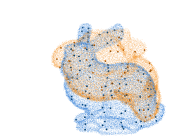


Despite decades of research, rigid point cloud registration is still an active and challenging problem since it has the chicken-and-egg property [LH07]. Specifically, the registration problem comprises two mutually interlocked sub-problems: pose and correspondence estimations. If one sub-problem is solved, another sub-problem will be solved accordingly. Commonly, existing registration methods are classified into two categories based on the requirement of correspondences or not, which are correspondence-based registration (e.g., fast global regis-

tration, FGR [ZPK16]) and simultaneous pose and correspondence registration (SPCR) (e.g., iterative closest point, ICP [BM92]). The widely used ICP is a local optimization method for SPCR, meaning it is highly dependent on the initialization of transformation and thus prone to fall into local minima, as shown in Fig. 4.1(d-2). The global methods for SPCR, however, deliver relatively low efficiency (e.g., globally optimal ICP, Go-ICP [Yan+16]), as shown in Fig. 4.1(d-3). Thus, the correspondence-based registration approaches are gradually attracting attention since they are initialization-free and more efficient [ZPK16; YSC20]. This chapter focuses on the correspondence-based registration problem, while, interestingly, the proposed approach can also be extended to solve the challenging SPCR problem.

Current 3D feature matching approaches have achieved satisfactory development. However, outlier correspondences are still inevitable either for handcrafted or learning-based descriptors [CPK19; Hua+21; Yan+22a]. Several paradigms have been extensively developed to implement robust registration, of which the *consensus maximization* (a.k.a. *inlier set maximization*) is inherently robust to outliers without smoothing or trimming to change the objective function [Li09; Cam+18]. Random sample consensus (RANSAC) [FB81] is the most popular heuristic method for solving the consensus maximization problem of correspondence-based registration. However, RANSAC and its variants are non-deterministic and only generate satisfactory solutions with a certain probability due to the random sampling mechanism [Le+19; BM21].

More recently, many global and deterministic methods based on the branch and bound (BnB) framework have been applied to solve the point cloud registration problem with optimality guarantees [HK09; Yan+16; CP16; Bus+16; Str+17; BC17; Liu+18b; Che+22c]. However, the computational complexity of BnB optimization is exponential to the dimensionality of the solution domain in the worst-case. Most studies address the issue by jointly searching for the optimal solution in  $\mathbb{SE}(3)$  [Yan+16; CP16; Bus+16]. In order to improve the algorithm efficiency, one direction is utilizing the known gravity directions measured by inertial measurement units (IMUs) to reduce the dimension of the parameter space to 4-dimensional [Cai+19]. Another direction for reducing the problem dimension is to decompose the original problem into two 3-DOF sub-problems by leveraging the geometric properties [Str+17; BC17; Liu+18b; YSC20]. Typically, two unique categories of features are employed for pose decoupling, i.e., the rotation invariant features (RIFs) [Liu+18b; Wan+21a] and the translation invariant measurements (TIMs) [Jia+22; YSC20]. Nonetheless, the pairwise features increase the number of input data quadratically, resulting in limited efficiency gains. Furthermore, a more efficient strategy is proposed based on the rotation decomposition, which decouples 6-DOF transformation into *i)* (2+1)-DOF, i.e., 2-DOF rotation axis and 1-DOF of translation along the axis, and *ii)* (1+2)-DOF, i.e., the remaining 1-DOF rotation and 2-DOF translation [Che+22c].

In this chapter, we propose an efficient and deterministic search strategy based on residual projections for the rigid registration problem, in which a novel pose decoupling strategy is introduced. Specifically, we decouple the 6-DOF original problem into three search sub-problems by projecting the residuals based on the *Chebyshev distance*, i.e.,  $L_\infty$  residual [SH06; KH08], on the coordinate axes. We then define the consensus maximization ob-

Correspondence-based (FPFH descriptor)	 (a-1) Initial 94.20% outlier - -	 (a-2) GORE [BC17] $E_R = 0.041^\circ$ $E_t = 0.055\text{m}$ $time = 403.1\text{s}$	 (a-3) TEASER [YSC20] $E_R = 0.069^\circ$ $E_t = 0.067\text{m}$ $time = 1.357\text{s}$	 (a-4) Ours $E_R = \mathbf{0.004^\circ}$ $E_t = \mathbf{0.042\text{m}}$ $time = \mathbf{0.307\text{s}}$
Correspondence-based (FPFH descriptor)	 (b-1) Initial 98.45% outlier - -	 (b-2) GORE [BC17] $E_R = \mathbf{0.055^\circ}$ $E_t = \mathbf{0.018\text{m}}$ $time = 687.7\text{s}$	 (b-3) TEASER [YSC20] $E_R = 0.197^\circ$ $E_t = 0.031\text{m}$ $time = 5.917\text{s}$	 (b-4) Ours $E_R = 0.104^\circ$ $E_t = \mathbf{0.018\text{m}}$ $time = \mathbf{1.436\text{s}}$
Correspondence-based (FCGF descriptor)	 (c-1) Initial 73.73% outlier - -	 (c-2) FGR [ZPK16] $E_R = 0.220^\circ$ $E_t = 0.521\text{m}$ $time = 1.258\text{s}$	 (c-3) GC-RANSAC [BM21] $E_R = 0.294^\circ$ $E_t = 0.510\text{m}$ $time = 1.576\text{s}$	 (c-4) Ours $E_R = \mathbf{0.212^\circ}$ $E_t = \mathbf{0.374\text{m}}$ $time = \mathbf{0.562\text{s}}$
Correspondence-free	 (d-1) Initial 60% overlap - -	 (d-2) ICP [BM92] $E_R = 173.7^\circ$ $E_t = 0.240\text{m}$ $time = \mathbf{0.069\text{s}}$	 (d-3) GoICP [Yan+16] $E_R = 2.092^\circ$ $E_t = 0.017\text{m}$ $time = 80.40\text{s}$	 (d-4) Ours $E_R = \mathbf{0.061^\circ}$ $E_t = \mathbf{0.004\text{m}}$ $time = 2.524\text{s}$

**Figure 4.1:** The proposed method can efficiently address the rigid registration problem in different scenarios with high outlier rates or low overlap rates. For the correspondence-based registration problem, the input correspondences are generated by the traditional descriptor FPFH [RBB09] and the learning-based descriptor FCGF [CPK19]. The input point clouds are selected from (a) Bremen dataset [BEN13], (b) ETH dataset [TWS14], (c) KITTI dataset [GLU12], and (d) Bunny dataset [CL96], respectively. The source point cloud is green, the target point cloud is yellow, and the aligned point cloud is blue. Compared with state-of-the-art (SOTA) correspondence-based methods, the proposed method achieves significant performance in terms of robustness and efficiency. Besides, the proposed method also can solve the SPCR problem efficiently and robustly.

jective function for each sub-problem and apply a BnB-based optimization method to search for the solution globally and deterministically while obtaining the consensus set. A novel polynomial-time upper bound is derived based on the *interval stabbing* technique [BC17; Cai+19; PTV22] for the proposed objective. The proposed BnB algorithm searches for three 2-DOF rotation matrix components individually. Meanwhile, three 1-DOF translation projections on the coordinate axes are implicitly estimated by interval stabbing. After solving these three sub-problems, we can obtain the solution to the 6-DOF registration problem, as well as the final consensus set. Finally, in order to get refined and valid results, the final rotation and translation are re-estimated by using Singular Value Decomposition (SVD) on the estimated

consensus set.

Contrary to existing methods that search in the three-dimensional domain via BnB, the proposed method allows searching only in the two-dimensional parameter space, thus enhancing the computational efficiency, as shown in Fig. 4.1. In addition, the proposed method requires no initialization of the translation domain, which is challenging to accurately determine in different practical scenarios. Therefore, it avoids the problems that would arise when the translation domain is not initialized correctly. Notably, we can also partially verify if the solution is valid by checking whether the coarse solutions of the three sub-problems are orthogonal before SVD. This is because rotation matrices are inherently orthogonal, with a determinant of 1.

The main contributions of this chapter can be summarized as follows:

- This chapter proposes a novel pose decoupling strategy based on the  $L_\infty$  residual projections. Compared with existing methods, our approach searches for the solution in a lower-dimensional parameter space, thereby improving search efficiency.
- This chapter a novel deterministic BnB-based search method for the decoupled sub-problems. The specific upper bound is derived based on the *interval stabbing* technique, allowing a further dimensionality reduction of the branching space.
- Due to its significant robustness, the proposed method can be extended to solve the challenging SPCR problem. The proposed upper bound is adapted to the SPCR objective by *interval merging* technique.

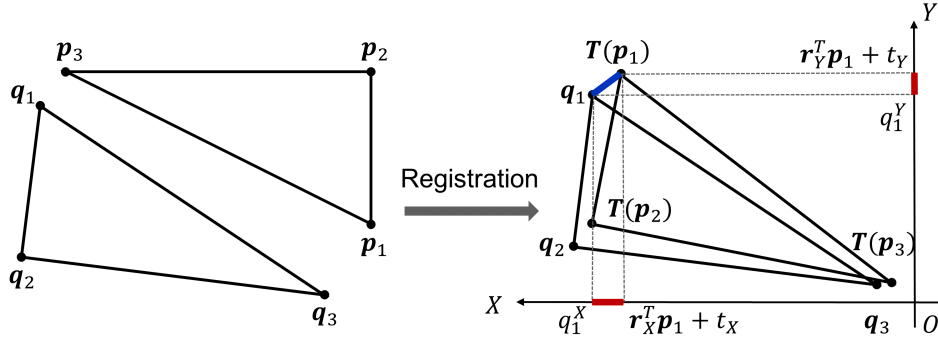
The rest of this chapter is organized as follows: Section 4.2 illustrates the problem formulation of the proposed method. Section 4.3 demonstrates the principle and details of our method. Section 4.4 presents extensive experimental results on both synthetic and real-world datasets. Finally, Section 4.5 gives a summary.

## 4.2 Problem Formulation

### 4.2.1 Inlier Set Maximization

Given the source point cloud  $\mathcal{P}$  and the target point cloud  $\mathcal{Q}$ , a set of putative correspondences  $\mathcal{K} = \{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^N$  is extracted by matching points between  $\mathcal{P}$  and  $\mathcal{Q}$ , where  $\mathbf{p}_i, \mathbf{q}_i \in \mathbb{R}^3$ , and  $N$  is the correspondences number. The proposed method aims to estimate the rigid transformation between the source and target point clouds. Specifically, the 6-DOF transformation matrix  $T \in \mathbb{SE}(3)$  is formed by the 3-DOF rotation matrix  $R \in \mathbb{SO}(3)$  and the 3-DOF translation vector  $\mathbf{t} \in \mathbb{R}^3$ . The rotation matrix  $R$  is an orthogonal matrix in which the columns and rows are orthogonal vectors, i.e.,  $RR^T = I$ , with a determinant of 1. Formally, the inlier set maximization formulation is adopted for the robust registration problem:

$$T^* = \arg \max_{T \in \mathbb{SE}(3)} E(T(\mathcal{P}), \mathcal{Q}), \quad (4.1)$$



**Figure 4.2:** A toy 2D registration example to demonstrate  $L_\infty$  residual projection. Specifically,  $\{(p_i, q_i)\}_{i=1}^3$  is the set of input correspondences,  $r_j$ ,  $j = X, Y$ , is the transpose of each row of the rotation matrix, and  $t_j$ ,  $j = X, Y$ , is the component of the translation vector. The red line segments represent the projections of the residual on the coordinate axes  $X$  and  $Y$ , i.e.,  $|r_j^T p_1 + t_j - q_1^j|$ ,  $j = X, Y$ . The inlier constraint for  $L_\infty$  residual indicates that  $(p_1, q_1)$  is an inlier only if both residual projections on the coordinate axes are not larger than the inlier threshold.

where  $E$  is the objective function for calculating the cardinality of the inlier set.

Different from existing approaches [Yan+16; Bus+16; BC17; Cai+19; Che+22c] that commonly employ the  $L_2$  residual to measure the alignment, the *Chebyshev distance* is applied, i.e.,  $L_\infty$  residual [SH06; KH08; Liu+18b], to build the robust objective function. Therefore, considering the presence of noise, we estimate the rotation and translation that maximize the objective:

$$E(\mathbf{R}, \mathbf{t} | \mathcal{K}, \epsilon) = \sum_{i=1}^N \mathbb{I}(\|\mathbf{R}p_i + \mathbf{t} - q_i\|_\infty \leq \epsilon), \quad (4.2)$$

where  $\mathbb{I}(\cdot)$  is the indicator function that returns 1 if the input condition is true and 0 otherwise,  $\|\cdot\|_\infty$  denotes the  $L_\infty$ -norm, and  $\epsilon$  is the inlier threshold.

## 4.2.2 Residual Projections and Pose Decoupling

Mathematically, we apply the following definitions to derive the residual projections. Firstly, we denote the rotation matrix as

$$\mathbf{R} \triangleq \begin{bmatrix} r_{X1} & r_{X2} & r_{X3} \\ r_{Y1} & r_{Y2} & r_{Y3} \\ r_{Z1} & r_{Z2} & r_{Z3} \end{bmatrix} = [\mathbf{r}_X \quad \mathbf{r}_Y \quad \mathbf{r}_Z]^T. \quad (4.3)$$

where  $\mathbf{r}_j = [r_{j1}, r_{j2}, r_{j3}]^T$ ,  $j = X, Y, Z$ , is the transpose of each row of the rotation matrix. The translation vector is

$$\mathbf{t} \triangleq [t_X, t_Y, t_Z]^T. \quad (4.4)$$

Given the definitions of  $\mathbf{R}$  and  $\mathbf{t}$ , according to the definition of Chebyshev distance, the

inlier constraint in the objective function (4.2) can be rewritten as

$$\|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|_\infty \leq \epsilon \quad (4.5a)$$

$$\Leftrightarrow \left\| \begin{bmatrix} \mathbf{r}_X^T \\ \mathbf{r}_Y^T \\ \mathbf{r}_Z^T \end{bmatrix} \mathbf{p}_i + \begin{bmatrix} t_X \\ t_Y \\ t_Z \end{bmatrix} - \begin{bmatrix} q_i^X \\ q_i^Y \\ q_i^Z \end{bmatrix} \right\|_\infty \leq \epsilon \quad (4.5b)$$

$$\Leftrightarrow \max \left\{ \begin{array}{l} |\mathbf{r}_X^T \mathbf{p}_i + t_X - q_i^X|, \\ |\mathbf{r}_Y^T \mathbf{p}_i + t_Y - q_i^Y|, \\ |\mathbf{r}_Z^T \mathbf{p}_i + t_Z - q_i^Z| \end{array} \right\} \leq \epsilon \quad (4.5c)$$

$$\Leftrightarrow \begin{cases} |\mathbf{r}_X^T \mathbf{p}_i + t_X - q_i^X| \leq \epsilon, \\ |\mathbf{r}_Y^T \mathbf{p}_i + t_Y - q_i^Y| \leq \epsilon, \\ |\mathbf{r}_Z^T \mathbf{p}_i + t_Z - q_i^Z| \leq \epsilon \end{cases} \quad (4.5d)$$

$$\Leftrightarrow \begin{cases} \mathbb{I}(|\mathbf{r}_X^T \mathbf{p}_i + t_X - q_i^X| \leq \epsilon) = 1, \\ \mathbb{I}(|\mathbf{r}_Y^T \mathbf{p}_i + t_Y - q_i^Y| \leq \epsilon) = 1, \\ \mathbb{I}(|\mathbf{r}_Z^T \mathbf{p}_i + t_Z - q_i^Z| \leq \epsilon) = 1 \end{cases} \quad (4.5e)$$

where  $\mathbf{q}_i \triangleq [q_i^X, q_i^Y, q_i^Z]^T$ , and  $|\mathbf{r}_j^T \mathbf{p}_i + t_j - q_i^j|$ ,  $j = X, Y, Z$ , are projections of the  $i$ -th residual on the coordinate axes, as shown in Fig. 4.2. Then we can set  $\mathbb{I}(|\mathbf{r}_j^T \mathbf{p}_i + t_j - q_i^j| \leq \epsilon) = \mathcal{L}_i^j$ . Therefore, the objective function (4.2) can be reformulated as

$$E(\mathbf{R}, \mathbf{t} | \mathcal{K}, \epsilon) = \sum_{i=1}^N \mathbb{I}(\mathcal{L}_i^X \wedge \mathcal{L}_i^Y \wedge \mathcal{L}_i^Z), \quad (4.6)$$

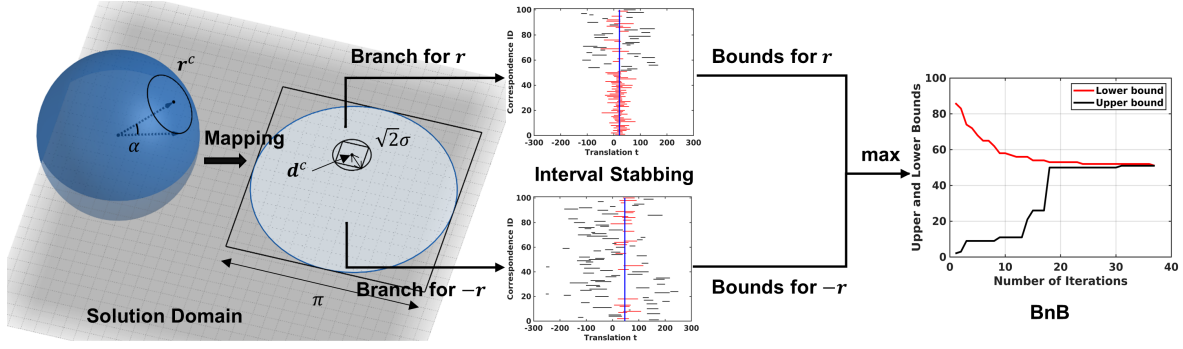
where  $\wedge$  is the logical AND operation.

Geometrically, the objective function (4.6) indicates that, given an arbitrary correspondence  $(\mathbf{p}_i, \mathbf{q}_i)$  and the inlier threshold  $\epsilon$ , only when the residual projections on the  $X$ ,  $Y$ , and  $Z$  coordinate axes are not larger than  $\epsilon$ ,  $(\mathbf{p}_i, \mathbf{q}_i)$  is an inlier, as shown in Fig. 4.2. Notably, these three conditions are equally independent. Accordingly, we may reduce the original constraint in Eq. (4.6) as three separate constraints, i.e.,  $\mathcal{L}_i^X$ ,  $\mathcal{L}_i^Y$ , and  $\mathcal{L}_i^Z$ .

In this way, the original search problem for the transformation in  $\mathbb{S}\mathbb{E}(3)$  can be decoupled into three sub-problems. The inlier set maximization objective for each sub-problem can be

$$E_j(\mathbf{r}_j, t_j | \mathcal{K}, \epsilon) = \sum_{i=1}^N \mathbb{I}(|\mathbf{r}_j^T \mathbf{p}_i + t_j - q_i^j| \leq \epsilon), j = X, Y, Z. \quad (4.7)$$

In other words, we reformulate the  $L_\infty$  residual-based objective function in the form of residual projections. Then we decompose the joint constraint into three independent constraints to decouple the original registration problem into three sub-problems, i.e.,  $\max E_X(\mathbf{r}_X, t_X | \mathcal{K}, \epsilon)$ ,  $\max E_Y(\mathbf{r}_Y, t_Y | \mathcal{K}, \epsilon)$ , and  $\max E_Z(\mathbf{r}_Z, t_Z | \mathcal{K}, \epsilon)$ . The following section will introduce a step-wise search strategy to solve these three sub-problems.



**Figure 4.3:** The solution domain before and after *exponential mapping*, and the pipeline of the proposed BnB algorithm. The original solution domain of the vector  $\mathbf{r}$  is a *unit sphere* in the 3D Euclidean space. The exponential mapping method maps the unit sphere to two identical 2D disks, representing the solution domains of  $\mathbf{r}$  and  $-\mathbf{r}$ , respectively. We can only branch one 2D-disk domain during each iteration, followed by the calculation of upper and lower bounds for each sub-branch. The proposed BnB algorithm converges until the optimal solution  $\mathbf{r}^*$  is found, and the optimal  $t^*$  is found by *interval stabbing* simultaneously. In the visualization results of interval stabbing, the black line segments are the candidate intervals of each correspondence, and the red line segments are the intervals crossed by the blue probe with the *max-stabbing number*. The probe position is the *max-stabbing position*.

### 4.3 Step-wise Search Strategy Based on Branch and Bound

Branch and bound (BnB) is an algorithm framework for global optimization. To design the BnB-based algorithm, two main aspects need to be addressed: *i)* how to parameterize and branch the solution domain, and *ii)* how to efficiently calculate the upper and lower bounds. Then, the BnB-based algorithm can recursively divide the solution domain into smaller spaces and prune the sub-branches by upper and lower bounds until convergence.

#### 4.3.1 Parametrization of Solution Domain

##### Rotation

For each sub-problem of the objective function (4.7), the unknown-but-sought vector  $\mathbf{r}_j$  (denoted by  $\mathbf{r}$  in this section) is on the surface of a *unit sphere* (denoted by  $\mathbb{S}^2$ ). Then we divide the unit sphere into two unit hemispheres ( $\mathbb{S}^{2+}$  and  $\mathbb{S}^{2-}$ ) to represent the parameter spaces of the “positive” vector  $\mathbf{r}$  and the “negative” vector  $-\mathbf{r}$ . The “upper” hemisphere is defined as

$$\mathbb{S}^{2+} = \{\mathbf{r} | \mathbf{r}^T \mathbf{r} = 1, r_3 \geq 0\}, \quad (4.8)$$

where  $\mathbf{r} \triangleq [r_1, r_2, r_3]^T$  is a unit vector in  $\mathbb{R}^3$ . Geometrically, since these two hemispheres are centrally symmetric, the “lower” hemisphere is  $\mathbb{S}^{2-}$  which can be seen as  $-\mathbb{S}^{2+}$ . In order to parametrize  $\mathbb{S}^{2+}$  and  $\mathbb{S}^{2-}$  minimally, we introduce the *exponential mapping* [LCK20b; Liu + 22] technique to map a 3-dimensional hemisphere to a 2-dimensional disk efficiently. Specifically, given a vector  $\mathbf{r} \in \mathbb{S}^{2+}$ , it can be represented by a corresponding point  $\mathbf{d} \in \mathbb{R}^2$  in the 2D disk, i.e.,

$$\mathbf{r}^T = [\sin(\gamma)\hat{\mathbf{d}}^T, \cos(\gamma)], \quad \text{and} \quad \mathbf{d} = \gamma\hat{\mathbf{d}} \quad (4.9)$$

where  $\gamma \in [0, \pi/2]$ , and  $\hat{\mathbf{d}}$  is a unit vector in  $\mathbb{R}^2$ . Notably, the range of  $\gamma$  corresponds to  $r_3 \geq 0$ , and its maximum corresponds to the radius of the 2D disk, i.e.,  $\pi/2$ , as shown in Fig. 4.3. For a vector  $-\mathbf{r} \in \mathbb{S}^{2-}$ , we define another exponential mapping method,

$$-\mathbf{r}^T = -[\sin(\gamma)\hat{\mathbf{d}}^T, \cos(\gamma)]. \quad (4.10)$$

Accordingly, the total solution domain (unit sphere) is mapped as two identical 2D disks, which represent the parameter spaces of  $\mathbf{r}$  and  $-\mathbf{r}$ , respectively. Compared to the unit sphere representation within three parameters and a unit-norm constraint, the exponential mapping is a more compact representation within only two parameters [Liu+22]. Meanwhile, for ease of operation, a circumscribed square of the disk domain is initialized as the domain of  $\mathbf{r}$  in the proposed BnB algorithm, and the domain of  $-\mathbf{r}$  is relaxed in the same way.

Further, we introduce the following lemma [LCK20b] about the exponential mapping between  $\mathbb{S}^{2+}$  and  $\mathbb{R}^2$ .

**Lemma 4.1.**  $\mathbf{r}_a, \mathbf{r}_b \in \mathbb{S}^{2+}$  are two vectors in the unit hemisphere, and  $\mathbf{d}_a, \mathbf{d}_b \in \mathbb{R}^2$  are corresponding points in the 2D disk. Then we have

$$\angle(\mathbf{r}_a, \mathbf{r}_b) \leq \|\mathbf{d}_a - \mathbf{d}_b\|. \quad (4.11)$$

According to Lemma 4.1, we can obtain the following proposition.

**Proposition 4.1.** Given a sub-branch of the square-shaped domain  $\mathbb{B}$ , its center is  $\mathbf{d}^c \in \mathbb{R}^2$  and half-side length is  $\sigma$ . For  $\forall \mathbf{d} \in \mathbb{B}$ , we have

$$\angle(\mathbf{r}, \mathbf{r}^c) \leq \|\mathbf{d} - \mathbf{d}^c\| \leq \sqrt{2}\sigma, \quad (4.12)$$

where  $\mathbf{r}$  and  $\mathbf{r}^c$  correspond to  $\mathbf{d}$  and  $\mathbf{d}^c$ , respectively.

Defining  $\alpha \triangleq \max \angle(\mathbf{r}, \mathbf{r}^c)$ , we can obtain  $\alpha \leq \sqrt{2}\sigma$  with Proposition 4.1, as shown in Fig. 4.3. Geometrically, Proposition 4.1 indicates that one square-shaped sub-branch of the 2D disk domain is relaxed to a spherical patch of the 3D unit sphere. In addition, Lemma 4.1 and Proposition 4.1 hold for both hemispheres  $\mathbb{S}^{2+}$  and  $\mathbb{S}^{2-}$ . In this study, we apply Proposition 4.1 as one of the fundamental parts to derive the proposed bound functions.

### Translation

Estimating the translation component  $t_j \in \mathbb{R}$ ,  $j = X, Y, Z$  in the objective function (4.7) is a 1-dimensional problem. The translation is unconstrained, and it is not easy to estimate a suitable solution domain accurately in advance for various practical scenarios. Existing BnB-based approaches [Yan+16; Che+22c; Liu+18b] commonly initialize the translation domain as a redundant space and search it exhaustively, leading to a significant decrease in efficiency. Meanwhile, if the translation domain is not initialized correctly, the algorithm may not find the optimal (correct) solution since the optimal solution may be excluded from the initial search domain.



This study proposes an *interval stabbing*-based method to estimate the translation components  $\{t_X, t_Y, t_Z\}$  without any prior information on the size of the translation domain, which can effectively reduce the total parameter space and improve the algorithm efficiency. It also avoids the problems that may arise when the translation initialization is incorrect. The proposed method will be described thoroughly in Section 4.3.2.

### 4.3.2 Interval Stabbing and Bounds

We first introduce the following lemma to derive the bounds for the objective function (4.7).

**Lemma 4.2.** *Given an arbitrary consensus maximization objective  $F(x|A) = \sum_{i=1}^M \mathcal{F}_i(x, a_i)$ , where  $x$  is the variable to be calculated,  $A = \{a_i\}_{i=1}^M$  is the set of input measurements, and  $\mathcal{F}_i(x, a_i)$  is an indicator function with a certain constraint. Then we have*

$$\max_x F(x|A) = \max_x \sum_{i=1}^M \mathcal{F}_i(x, a_i) \leq \sum_{i=1}^M \max_x \mathcal{F}_i(x, a_i). \quad (4.13)$$

*Proof.* For the  $i$ -th input measurement  $a_i$ , we can obtain  $\mathcal{F}_i(x, a_i) \leq \max_x \mathcal{F}_i(x, a_i) \leq 1$ . Therefore, it is obvious that the maximum of  $\sum_{i=1}^M \mathcal{F}_i(x, a_i)$  is not bigger than the sum of  $\max_x \mathcal{F}_i(x, a_i)$ .  $\square$

In this study, the upper and lower bounds are proposed as follows:

**Proposition 4.2** (Upper bound for  $\mathbb{S}^{2+}$ ). *Given a sub-branch of the square-shaped domain  $\mathbb{B}$ , whose center is  $\mathbf{d}^c \in \mathbb{R}^2$  (corresponds to  $\mathbf{r}_j^c \in \mathbb{S}^{2+}$ ) and half-side length is  $\sigma$ , the upper bound can be set as*

$$\bar{E}_j^+(\mathbb{B}) = \max_{t_j} \sum_{i=1}^N \mathbb{I}(t_j \in [t_j^{i-}, t_j^{i+}]), \quad (4.14a)$$

$$t_j^{i-} = -\epsilon - \|\mathbf{p}_i\| \cos\left(\max\left\{\angle\left(\mathbf{r}_j^c, \mathbf{p}_i\right) - \sqrt{2}\sigma, 0\right\}\right) + q_i^j, \quad (4.14b)$$

$$t_j^{i+} = \epsilon - \|\mathbf{p}_i\| \cos\left(\min\left\{\angle\left(\mathbf{r}_j^c, \mathbf{p}_i\right) + \sqrt{2}\sigma, \pi\right\}\right) + q_i^j. \quad (4.14c)$$

*Proof.* First, we rewrite the maximum of the objective function (4.7) as,

$$\max_{r_j, t_j} \sum_{i=1}^N \mathcal{L}_i^j = \max_{t_j} \max_{r_j} \sum_{i=1}^N \mathcal{L}_i^j. \quad (4.15)$$

Therefore, according to Lemma 4.2, we have

$$\max_{t_j} \max_{r_j} \sum_{i=1}^N \mathcal{L}_i^j \leq \max_{t_j} \sum_{i=1}^N \max_{r_j} \mathbb{I}\left(\left|\mathbf{r}_j^T \mathbf{p}_i + t_j - q_i^j\right| \leq \epsilon\right). \quad (4.16)$$

Additionally, given a sub-branch  $\mathbb{B}$ , according to the triangle inequality in spherical geom-

entry [Cam+18] and Proposition 4.1, we have

$$\angle(\mathbf{r}_j, \mathbf{p}_i) \leq \angle(\mathbf{r}_j^c, \mathbf{p}_i) + \angle(\mathbf{r}_j^c, \mathbf{r}_j) \quad (4.17a)$$

$$\leq \angle(\mathbf{r}_j^c, \mathbf{p}_i) + \alpha \quad (4.17b)$$

$$\leq \angle(\mathbf{r}_j^c, \mathbf{p}_i) + \sqrt{2}\sigma, \quad (4.17c)$$

and

$$\angle(\mathbf{r}_j, \mathbf{p}_i) \geq \angle(\mathbf{r}_j^c, \mathbf{p}_i) - \angle(\mathbf{r}_j^c, \mathbf{r}_j) \quad (4.18a)$$

$$\geq \angle(\mathbf{r}_j^c, \mathbf{p}_i) - \alpha \quad (4.18b)$$

$$\geq \angle(\mathbf{r}_j^c, \mathbf{p}_i) - \sqrt{2}\sigma. \quad (4.18c)$$

Thus, according to  $\mathbf{r}_j^T \mathbf{p}_i = \|\mathbf{r}_j\| \|\mathbf{p}_i\| \cos \angle(\mathbf{r}_j, \mathbf{p}_i)$  and  $\|\mathbf{r}_j\| = 1$ , we have

$$\begin{aligned} \mathbf{r}_j^T \mathbf{p}_i \in & \left[ \|\mathbf{p}_i\| \cos \left( \min \left\{ \angle(\mathbf{r}_j^c, \mathbf{p}_i) + \sqrt{2}\sigma, \pi \right\} \right), \right. \\ & \left. \|\mathbf{p}_i\| \cos \left( \max \left\{ \angle(\mathbf{r}_j^c, \mathbf{p}_i) - \sqrt{2}\sigma, 0 \right\} \right) \right]. \end{aligned} \quad (4.19)$$

Then, given a sub-branch  $\mathbb{B}$ , whose center is  $\mathbf{d}^c$  (corresponds to  $\mathbf{r}_j^c$ ) and half-side length is  $\sigma$ , we can observe that,

$$\max_{\mathbf{r}_j} \mathbb{I} \left( \left| \mathbf{r}_j^T \mathbf{p}_i + t_j - q_i^j \right| \leq \epsilon \right) \quad (4.20a)$$

$$= \max_{\mathbf{r}_j} \mathbb{I} \left( -\epsilon - \mathbf{r}_j^T \mathbf{p}_i + q_i^j \leq t_j \leq \epsilon - \mathbf{r}_j^T \mathbf{p}_i + q_i^j \right) \quad (4.20b)$$

$$\leq \mathbb{I} \left( t_j \in \left[ t_j^{i-}, t_j^{i+} \right] \right), \quad (4.20c)$$

where

$$t_j^{i-} = -\epsilon - \|\mathbf{p}_i\| \cos \left( \max \left\{ \angle(\mathbf{r}_j^c, \mathbf{p}_i) - \sqrt{2}\sigma, 0 \right\} \right) + q_i^j, \quad (4.21a)$$

$$t_j^{i+} = \epsilon - \|\mathbf{p}_i\| \cos \left( \min \left\{ \angle(\mathbf{r}_j^c, \mathbf{p}_i) + \sqrt{2}\sigma, \pi \right\} \right) + q_i^j. \quad (4.21b)$$

Then,

$$\sum_{i=1}^N \max_{\mathbf{r}_j} \mathbb{I} \left( \left| \mathbf{r}_j^T \mathbf{p}_i + t_j - q_i^j \right| \leq \epsilon \right) \leq \sum_{i=1}^N \mathbb{I} \left( t_j \in \left[ t_j^{i-}, t_j^{i+} \right] \right) \quad (4.22)$$

Finally, we have

$$\max_{\mathbf{r}_j, t_j} \sum_{i=1}^N \mathcal{L}_i^j \leq \max_{t_j} \sum_{i=1}^N \mathbb{I} \left( t_j \in \left[ t_j^{i-}, t_j^{i+} \right] \right) \quad (4.23)$$

Therefore, Proposition 4.2 is proved.  $\square$

**Proposition 4.3** (Upper bound for  $\mathbb{S}^{2-}$ ). *Given a sub-branch of the square-shaped domain  $\mathbb{B}$ , whose center is  $\mathbf{d}^c \in \mathbb{R}^2$  (corresponds to  $-\mathbf{r}_j^c \in \mathbb{S}^{2-}$ ) and half-side length is  $\sigma$ , the upper bound*

can be set as

$$\overline{E}_j(\mathbb{B}) = \max_{t_j} \sum_{i=1}^N \mathbb{I}(t_j \in [t_j^{i-}, t_j^{i+}]), \quad (4.24a)$$

$$t_j^{i-} = -\epsilon + \|\mathbf{p}_i\| \cos\left(\min\left\{\angle\left(\mathbf{r}_j^c, \mathbf{p}_i\right) + \sqrt{2}\sigma, \pi\right\}\right) + q_i^j, \quad (4.24b)$$

$$t_j^{i+} = \epsilon + \|\mathbf{p}_i\| \cos\left(\max\left\{\angle\left(\mathbf{r}_j^c, \mathbf{p}_i\right) - \sqrt{2}\sigma, 0\right\}\right) + q_i^j. \quad (4.24c)$$

*Proof.* The proof is similar to Proposition 4.2, which is simple enough that we omit it.  $\square$

Although the upper bounds in Proposition 4.2 and Proposition 4.3 are theoretically provided, we still need to find an appropriate method to compute them. Mathematically, the calculation of the upper bounds is a typical *interval stabbing* problem [De +97]. As shown in Fig. 4.3, the interval stabbing problem aims to find a probe (i.e., the blue line segment) that stabs the maximum number of intervals. There has been a deterministic and polynomial-time algorithm [Cai+19] to solve the interval stabbing problem. More details are given in [De +97; Cai+19].

By utilizing the interval stabbing technique to compute the upper bounds, the proposed BnB-based method only needs to search a 2-dimensional solution domain, thereby improving the algorithm efficiency. Meanwhile, the translation projections  $\{t_X, t_Y, t_Z\}$  are implicitly estimated by interval stabbing without requiring the initialization of the translation domain. In other words, the interval stabbing approach returns not only the *max-stabbing number* (i.e., the upper bound), but also the *max-stabbing position* (i.e., the estimation of  $t_j$ ).

To sum up, considering the total solution domain  $\mathbb{S}^{2+}$  and  $\mathbb{S}^{2-}$ , we have the following proposition.

**Proposition 4.4** (Upper bound for  $\mathbb{S}^2$ ). *Given a sub-branch of the square-shaped domain  $\mathbb{B}$ , whose center is  $\mathbf{d}^c \in \mathbb{R}^2$  and half-side length is  $\sigma$ , the upper bound of the objective function (4.7), can be set as*

$$\overline{E}_j(\mathbb{B}) = \max\{\overline{E}_j^+(\mathbb{B}), \overline{E}_j^-(\mathbb{B})\}. \quad (4.25)$$

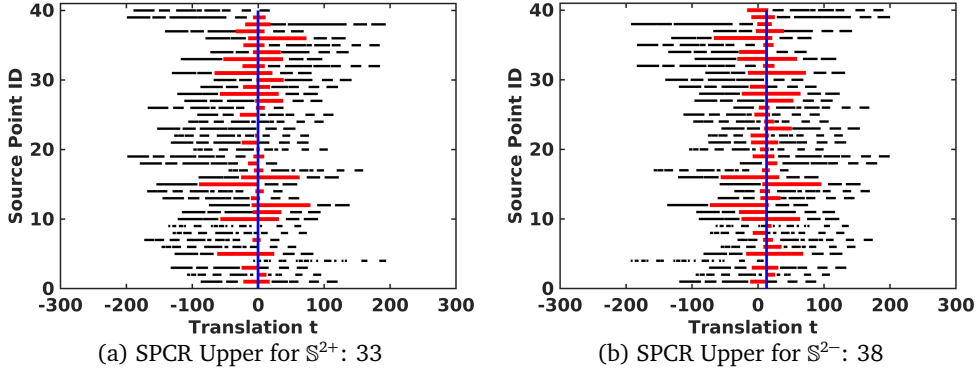
*Proof.* The maximum of these two upper bounds is not smaller than the maximum of the objective function (4.7). Therefore,  $\overline{E}_j(\mathbb{B})$  is the final upper bound of the objective function (4.7).  $\square$

**Proposition 4.5** (Lower bound for  $\mathbb{S}^2$ ). *Given a sub-branch of the square-shaped domain  $\mathbb{B}$ , whose center is  $\mathbf{d}^c \in \mathbb{R}^2$  and half-side length is  $\sigma$ , the lower bound of the objective function (4.7) can be set as*

$$\underline{E}_j(\mathbb{B}) = \max\{\underline{E}_j^+(\mathbb{B}), \underline{E}_j^-(\mathbb{B})\}, \quad (4.26a)$$

$$\underline{E}_j^+(\mathbb{B}) = \sum_{i=1}^N \mathbb{I}\left(\left|(\mathbf{r}_j^c)^T \mathbf{p}_i + \overline{t}_j^+ - q_i^j\right| \leq \epsilon\right), \quad (4.26b)$$

$$\underline{E}_j^-(\mathbb{B}) = \sum_{i=1}^N \mathbb{I}\left(\left|-(\mathbf{r}_j^c)^T \mathbf{p}_i + \overline{t}_j^- - q_i^j\right| \leq \epsilon\right), \quad (4.26c)$$



**Figure 4.4:** The visualization results for the 24-th iteration of a representative SPCR test on synthetic data. The black line segments are the intervals after *interval merging*, and the red line segments are the intervals crossed by the blue probe with the max-stabbing number. After interval merging, interval stabbing is utilized to calculate the upper bounds. Notably, the blue probe resulting from interval stabbing can only penetrate at most one interval for each point  $p_i$ . The final SPCR upper bound for  $\mathbb{S}^2$  is 38.

where  $\bar{t}_j^+$  is the max-stabbing position of the upper bound for  $\mathbb{S}^{2+}$ , and  $\bar{t}_j^-$  is the max-stabbing position of the upper bound for  $\mathbb{S}^{2-}$ .

*Proof.* The maximum of the objective function in the given sub-branch  $\mathbb{B}$  should be no less than any objective value at a specific point. Therefore,  $\underline{E}_j(\mathbb{B})$  is the lower bound of the objective function (4.7).  $\square$

**Proposition 4.6** (Bound convergence). *When the sub-branch of the square-shaped domain  $\mathbb{B}$  collapses to a single point whose center is  $\mathbf{d}^c \in \mathbb{R}^2$  and half-side length is zero, we can have*

$$\bar{E}_j(\mathbb{B}) = \underline{E}_j(\mathbb{B}). \quad (4.27)$$

*Proof.* When the sub-branch  $\mathbb{B}$  collapses to a single point, we have  $\sigma = 0$ , and the upper bound for  $\mathbb{S}^{2+}$  is

$$\bar{E}_j^+(\mathbb{B}) = \max_{t_j} \sum_{i=1}^N \mathbb{I}(t_j \in [t_j^{i-}, t_j^{i+}]), \quad (4.28a)$$

$$t_j^{i-} = -\epsilon - (\mathbf{r}_j^c)^\top \mathbf{p}_i + q_i^j, \quad (4.28b)$$

$$t_j^{i+} = \epsilon - (\mathbf{r}_j^c)^\top \mathbf{p}_i + q_i^j. \quad (4.28c)$$

The lower bound for  $\mathbb{S}^{2+}$  is

$$\underline{E}_j^+(\mathbb{B}) = \sum_{i=1}^N \mathbb{I}(|(\mathbf{r}_j^c)^\top \mathbf{p}_i + \bar{t}_j^+ - q_i^j| \leq \epsilon), \quad (4.29)$$

where  $\bar{t}_j^+$  is the max-stabbing position of the upper bound for  $\mathbb{S}^{2+}$ . Therefore,  $\bar{E}_j^+(\mathbb{B}) = \underline{E}_j^+(\mathbb{B})$ . Similarly, we can have  $\bar{E}_j^-(\mathbb{B}) = \underline{E}_j^-(\mathbb{B})$ . Finally,  $\bar{E}_j(\mathbb{B}) = \underline{E}_j(\mathbb{B})$  is proved.  $\square$

Based on the upper and lower bounds in Proposition 4.4 and Proposition 4.5, the proposed 2D BnB search algorithm for solving decoupled sub-problems is outlined in Algorithm 2. We

**Algorithm 2:** 2D BnB search for decoupled sub-problems

**Input:** The set of correspondences  $\mathcal{K} = \{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^N$ , inlier threshold  $\epsilon$ .

**Output:** Optimal solution  $\mathbf{r}_j^* \in \mathbb{S}^2$  and  $t_j^* \in \mathbb{R}$ .

- 1 Initialize the solution domain  $\mathbb{B}_0$ ;
- 2 Initialize the list of sub-branches  $\xi = \{\mathbb{B}_0\}$ ;
- 3 Initialize the lower bound  $L = 0$  and the upper bound  $U = N$ ;
- 4 Define function  $\delta(\mathbb{B})$  returns the center of sub-branch  $\mathbb{B}$ ;
- 5 Define function  $\eta(\mathbb{B})$  returns the max-stabbing position of sub-branch  $\mathbb{B}$  corresponding to the upper bound;
- 6 **while**  $U - L > 0$  **do**
- 7     Select a sub-branch  $\mathbb{B}$  with the maximal upper bound from  $\xi$ , i.e.,  
 $\mathbb{B} = \arg \max \bar{E}_j(\mathbb{B}_k), \mathbb{B}_k \in \xi$ ;
- 8     Subdivide  $\mathbb{B}$  into four sub-branches  $\{\mathbb{B}_1, \dots, \mathbb{B}_4\}$ ;
- 9     Insert  $\{\mathbb{B}_1, \dots, \mathbb{B}_4\}$  into  $\xi$  and eliminate  $\mathbb{B}$  from  $\xi$ ;
- 10    Update  $U = \max \bar{E}_j(\mathbb{B}_k), \mathbb{B}_k \in \xi$ ;
- 11    Update  $L = \max \{L, \underline{E}_j(\mathbb{B}_k)\}$  with  $\mathbb{B}_k \in \xi$ , if  $\underline{E}_j(\mathbb{B}_k) > L$ , set  $\mathbf{r}_j^* = \delta(\mathbb{B}_k)$  and  
 $t_j^* = \eta(\mathbb{B}_k)$ ;
- 12    Eliminate  $\mathbb{B}_k$  from  $\xi$  if  $\bar{E}_j(\mathbb{B}_k) < L, \mathbb{B}_k \in \xi$ ;
- 13 **end**

employ the depth-first search strategy [MSS08] to implement the proposed BnB algorithm. As we indicated in Section 4.3.1, although the initial solution domain  $\mathbb{S}^2$  is mapped to two identical 2D-disks, only one disk domain is branched, since the bounds of  $\mathbb{S}^{2+}$  and  $\mathbb{S}^{2-}$  can be computed separately in the same disk domain, as shown in Fig. 4.3. During each iteration, the branch with maximal upper bound is partitioned into four sub-branches since the current parameter space is only 2-dimensional. Then, the branch list is updated, and the upper and lower bounds for each sub-branch are estimated. The sub-branches that do not have a better solution than the best-so-far solution are eliminated. As the number of iterations increases, the gap between the upper and lower bounds gradually decreases. Until the gap reduces to zero, the proposed BnB algorithm obtains the optimal solutions  $(\mathbf{r}_j^*, t_j^*)$  and consensus set. As shown in [Str+17; Liu+18b; Che+22c], existing methods usually solve sub-problems sequentially. However, we can solve three sub-problems in an arbitrary order with Algorithm 2 and obtain the final results  $\mathbf{R}^*$  and  $\mathbf{t}^*$  by SVD refinement.

### 4.3.3 Simultaneous Pose and Correspondence Registration

This section extends the proposed correspondence-based registration method to address the challenging simultaneous pose and correspondence registration (SPCR) problem. The SPCR problem inherently poses greater complexity than the correspondence-based problem. Formally, given the source point cloud  $\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^M$  and the target point cloud  $\mathcal{Q} = \{\mathbf{q}_k\}_{k=1}^N$ , there are  $M \times N$  candidate correspondences totally. Similar to [Cam+18; Wan+21a], we define

the inlier set maximization objective function for the SPCR problem as

$$S_j(\mathbf{r}_j, t_j | \mathcal{P}, \mathcal{Q}, \epsilon) = \sum_{i=1}^M \max_k \mathbb{I} \left( \left| \mathbf{r}_j^T \mathbf{p}_i + t_j - q_k^j \right| \leq \epsilon \right), \quad (4.30)$$

where  $j = X, Y, Z$ . This formulation implies that for each point  $\mathbf{p}_i$ , as long as a sufficiently close point  $\mathbf{q}_k$  exists, then it contributes a maximum of 1 to the objective function. In other words, for each point  $\mathbf{p}_i$ , we may not need to explicitly check all  $N$  points in  $\mathcal{Q}$ .

The upper and lower bounds for the SPCR objective (4.30) are slightly different from those of correspondence-based registration, given by the following propositions. In addition, the optimization of objective (4.30) is also based on BnB.

**Proposition 4.7** (SPCR Upper bound for  $\mathbb{S}^2$ ). *Given a sub-branch of the square-shaped domain  $\mathbb{B}$ , whose center is  $\mathbf{d}^c \in \mathbb{R}^2$  and half-side length is  $\sigma$ , the SPCR upper bound for  $\mathbb{S}^{2+}$  can be set as*

$$\overline{S}_j^+(\mathbb{B}) = \max_{t_j} \sum_{i=1}^M \max_k \mathbb{I} \left( t_j \in [t_j^{ik-}, t_j^{ik+}] \right), \quad (4.31a)$$

$$t_j^{ik-} = -\epsilon - \|\mathbf{p}_i\| \cos \left( \max \left\{ \angle \left( \mathbf{r}_j^c, \mathbf{p}_i \right) - \sqrt{2}\sigma, 0 \right\} \right) + q_k^j, \quad (4.31b)$$

$$t_j^{ik+} = \epsilon - \|\mathbf{p}_i\| \cos \left( \min \left\{ \angle \left( \mathbf{r}_j^c, \mathbf{p}_i \right) + \sqrt{2}\sigma, \pi \right\} \right) + q_k^j. \quad (4.31c)$$

The SPCR upper bound for  $\mathbb{S}^{2-}$  can be set as

$$\overline{S}_j^-(\mathbb{B}) = \max_{t_j} \sum_{i=1}^M \max_k \mathbb{I} \left( t_j \in [t_j^{ik-}, t_j^{ik+}] \right), \quad (4.32a)$$

$$t_j^{ik-} = -\epsilon + \|\mathbf{p}_i\| \cos \left( \min \left\{ \angle \left( \mathbf{r}_j^c, \mathbf{p}_i \right) + \sqrt{2}\sigma, \pi \right\} \right) + q_k^j, \quad (4.32b)$$

$$t_j^{ik+} = \epsilon + \|\mathbf{p}_i\| \cos \left( \max \left\{ \angle \left( \mathbf{r}_j^c, \mathbf{p}_i \right) - \sqrt{2}\sigma, 0 \right\} \right) + q_k^j. \quad (4.32c)$$

The final SPCR upper bound for  $\mathbb{S}^2$  can be set as

$$\overline{S}_j(\mathbb{B}) = \max \left\{ \overline{S}_j^+(\mathbb{B}), \overline{S}_j^-(\mathbb{B}) \right\}. \quad (4.33)$$

*Proof.* The proof is similar to the proofs of Proposition 4.2, Proposition 4.3, and Proposition 4.4, hence we omit it.  $\square$

In Proposition 4.7, the challenge lies in finding a solution to the problem of  $\max_k \mathbb{I} \left( t_j \in [t_j^{ik-}, t_j^{ik+}] \right)$  while ensuring its maximum value does not exceed 1. Given that there are  $N$  intervals for each point  $\mathbf{p}_i$ , and these intervals may overlap, directly applying interval stabbing to these  $M \times N$  intervals is unfeasible (otherwise, the upper bound function value would be much larger than the maximum value of the objective function  $M$ ). Therefore, the *interval merging* technique [De +97] can be employed as a pre-processing method before applying the interval stabbing algorithm to calculate bounds (4.31a) and (4.32a). After interval merging, the max-stabbing probe can penetrate at most one interval for each point  $\mathbf{p}_i$ , meaning each point  $\mathbf{p}_i$  contributes a maximum of 1 to the upper bound functions (4.31a) and (4.32a). The complete interval merging algorithm for all points  $\{\mathbf{p}_i\}_{i=1}^M$  is presented in Algorithm 3.

Specifically, interval merging is executed one time for each point  $\mathbf{p}_i$ , and then a total of  $M$  times for point cloud  $\mathcal{P}$ . Subsequently, interval stabbing is employed on the merged intervals to compute upper bounds. An example of the visualization results of interval merging and stabbing is given in Fig. 4.4. Similarly, when computing the SPCR lower bound, we employ another indicator function to solve this “multi-interval” problem, as shown in Eq. (4.34b) and Eq. (4.34d) of the following Proposition 4.8.

**Proposition 4.8** (SPCR Lower bound for  $\mathbb{S}^2$ ). *Given a sub-branch of the square-shaped domain  $\mathbb{B}$ , whose center is  $\mathbf{d}^c \in \mathbb{R}^2$  and half-side length is  $\sigma$ , the SPCR lower bound can be set as*

$$\underline{S}_j(\mathbb{B}) = \max \left\{ \underline{S}_j^+(\mathbb{B}), \underline{S}_j^-(\mathbb{B}) \right\}, \quad (4.34a)$$

$$\underline{S}_j^+(\mathbb{B}) = \sum_{i=1}^M \mathbb{I}(\mathcal{M}_i^{j+} > 0), \quad (4.34b)$$

$$\mathcal{M}_i^{j+} = \sum_{k=1}^N \mathbb{I} \left( \left| (\mathbf{r}_j^c)^T \mathbf{p}_i + \bar{t}_j^+ - q_k^j \right| \leq \epsilon \right), \quad (4.34c)$$

$$\underline{S}_j^-(\mathbb{B}) = \sum_{i=1}^M \mathbb{I}(\mathcal{M}_i^{j-} > 0), \quad (4.34d)$$

$$\mathcal{M}_i^{j-} = \sum_{k=1}^N \mathbb{I} \left( \left| -(\mathbf{r}_j^c)^T \mathbf{p}_i + \bar{t}_j^- - q_k^j \right| \leq \epsilon \right), \quad (4.34e)$$

where  $\bar{t}_j^+$  is the max-stabbing position of the SPCR upper bound for  $\mathbb{S}^{2+}$ , and  $\bar{t}_j^-$  is the max-stabbing position of the SPCR upper bound for  $\mathbb{S}^{2-}$ .

*Proof.* The proof is similar to the proof of Proposition 4.5, hence we omit it.  $\square$

**Proposition 4.9** (SPCR bound convergence). *When the sub-branch of the square-shaped domain  $\mathbb{B}$  collapses to a single point whose center is  $\mathbf{d}^c \in \mathbb{R}^2$  and half-side length is zero, we can have*

$$\bar{S}_j(\mathbb{B}) = \underline{S}_j(\mathbb{B}). \quad (4.35)$$

*Proof.* The proof is similar to the proof of Proposition 4.6, hence we omit it.  $\square$

To improve the total efficiency, we can only solve the first sub-problem (e.g., maximize  $S_X(\mathbf{r}_X, t_X | \mathcal{P}, \mathcal{Q}, \epsilon)$ ) using the extended BnB-based SPCR approach. Then we can solve the second sub-problem (e.g., maximize  $E_Y(\mathbf{r}_Y, t_Y | \mathcal{K}, \epsilon)$ ) and third sub-problem (e.g., maximize  $E_Z(\mathbf{r}_Z, t_Z | \mathcal{K}, \epsilon)$ ) by Algorithm 2. This is because we can obtain the candidate inlier correspondences after solving the first SPCR sub-problem, which is implicitly determined by the residual projection constraint. Notably, partial outliers occasionally satisfy this constraint and cannot be removed. However, the proposed correspondence-based method can be applied to robustly address the two remaining sub-problems and generate the final consensus set.

**Algorithm 3:** Interval merging for SPCR bounds calculation

---

**Input:** Intervals  $\left\{ \left\{ \left[ t_j^{ik-}, t_j^{ik+} \right] \right\}_{i=1}^M \right\}_{k=1}^N$ .

**Output:** Merged intervals  $\psi = \{[a_l, b_l]\}_{l=1}^{N'}$ , where  $N' = \sum_{i=1}^M N_i$ .

- 1 Initialize the index  $i = 1$ ;
- 2 Initialize the list of merged intervals  $\psi = \emptyset$ ;
- 3 **while**  $i \leq M$  **do**
- 4     Sort the intervals  $\left\{ \left[ t_j^{ik-}, t_j^{ik+} \right] \right\}_{k=1}^N$  by left side  $t_j^{ik-}$  in ascending order;
- 5     Initialize the index  $k = 1, l = 1$ ;
- 6     Initialize  $a_1 = t_j^{i1-}, b_1 = t_j^{i1+}$ ;
- 7     **while**  $k < N$  **do**
- 8         **if**  $t_j^{ik+} \geq t_j^{i(k+1)-}$  **then**
- 9              $b_l = \max \{ t_j^{ik+}, t_j^{i(k+1)+} \}$ ;
- 10         **else**
- 11              $l = l + 1$ ;
- 12              $a_l = t_j^{i(k+1)-}, b_l = t_j^{i(k+1)+}$ ;
- 13         **end**
- 14          $k = k + 1$ ;
- 15     **end**
- 16     Insert the merged intervals into list  $\psi$ , i.e.,  $\psi = \psi \cup \{[a_l, b_l]\}_{l=1}^{N_i}$ , where  $N_i$  is the number of merged intervals for  $p_i$ ;
- 17 **end**

---

## 4.4 LiDAR Registration Experiments

This section compares the proposed method with SOTA correspondence-based methods on both synthetic and real-world datasets. Additionally, we evaluate the extended method against existing SPCR methods, specifically on synthetic data. We implement the proposed method in Matlab 2019b and conduct all experiments on a laptop with an i7-9750H CPU and 16GB RAM.

### 4.4.1 Experimental Setting

We denote the proposed method as **Ours**. The compared methods for correspondence-based registration are as follows,

- GORE [BC17]: A guaranteed outlier removal registration method based on BnB and pose decoupling. It is implemented in C++.
- RANSAC [FB81]: A typical consensus maximization registration approach implemented in Matlab. The maximum number of iterations is set to  $10^4$ .



- TEASER [YSC20]: A certifiable decoupling-based registration method with a robust cost function. It is implemented in C++.
- FGR [ZPK16]: A fast registration method with a robust cost function. It is implemented in C++.
- GC-RANSAC [BM21]: A variant of RANSAC-based registration method with improvements in local optimization. It is implemented in C++, and the maximum number of iterations is set to  $10^4$ .
- TR-DE [Che+22c]: A deterministic point cloud registration method based on BnB and pose decoupling. It is implemented in C++.
- DGR [CDK20]: A learning-based outlier rejection method employing the sparse convolutional network. It is implemented in Python.
- DHVR [Lee+21]: A learning-based outlier rejection method leveraging the Hough voting. It is implemented in Python.
- PointDSC [Bai+21]: A learning-based outlier rejection method utilizing the spatial consistency. It is implemented in Python.

Besides, the compared methods for SPCR are as follows,

- GO-ICP [Yan+16]: A 6-DOF global optimal registration method based on BnB. It is implemented in C++.
- GO-ICPT [Yan+16]: A variant of GO-ICP with outlier trimming.
- ICP [BM92]: A typical EM-type method implemented by *pcregistericp* function in MATLAB.
- CPD [MS10]: A robust GMM-based registration approach implemented in C.
- GMMReg [JV10]: A robust and general GMM-based registration method implemented in C.

Similar to [YSC20; Che+22c; Bai+21], the evaluation metrics for point cloud registration in this study include 1) rotation error  $E_R$ , 2) translation error  $E_t$ , 3) running time, 4) success rate  $SR$ , and 5)  $F1$ -score. The error definitions are as follows:

$$E_R = \arccos\left(\frac{\text{Tr}(\mathbf{R}_{gt}^{-1}\mathbf{R}^*) - 1}{2}\right), \quad (4.36a)$$

$$E_t = \|\mathbf{t}_{gt} - \mathbf{t}^*\|, \quad (4.36b)$$

where  $\mathbf{t}_{gt}$  and  $\mathbf{R}_{gt}$  are the ground truth,  $\mathbf{t}^*$  and  $\mathbf{R}^*$  are the estimated solutions, and  $\text{Tr}(\cdot)$  is the trace of a matrix. The successful cases must satisfy the predefined threshold for  $E_R$  and  $E_t$ . Besides, the definition of  $F1$ -score is given in [Bai+21].

**Table 4.1:** Controlled experiments with the extremely high number of correspondences. The results include average rotation error (°) | average translation error (m) | average running time (s). Bolded and underlined fonts indicate the first two best values.

Method	10k	20k	50k	100k	200k	500k
GORE	>1 hour					
TEASER	out-of-memory					
RANSAC	0.121 0.152 4.506	0.086 0.185 6.732	0.073 0.153 12.51	0.079 0.137 30.13	0.080 0.143 92.37	4.961 4.883 315.7
GC-RANSAC	0.111 0.176 2.636	0.100 0.200 9.976	122.4 141.3 20.64	139.3 142.0 20.64	126.3 138.4 20.64	133.4 148.5 20.64
FGR [ZPK16]	<u>0.021</u>   <b>0.010</b>   <u>1.540</u>	0.024  <b>0.013</b>   <u>2.477</u>	0.037  <b>0.022</b>   <u>6.346</u>	<u>0.031</u>   <b>0.018</b>   <u>12.13</u>	<u>0.024</u>   <b>0.018</b>   <u>23.97</u>	<u>0.047</u>   <b>0.025</b>   <u>68.74</u>
Ours	<b>0.016</b>   <u>0.017</u>   <b>0.397</b>	<b>0.022</b>   <u>0.028</u>   <b>0.649</b>	<b>0.025</b>   <u>0.025</u>   <b>1.649</b>	<b>0.025</b>   <u>0.028</u>   <b>2.963</b>	<b>0.023</b>   <u>0.027</u>   <b>6.111</b>	<b>0.018</b>   <b>0.025</b>   <b>15.56</b>

#### 4.4.2 Synthetic Data Experiments

In this section, we conduct various experiments on synthetic data to compare the performance of the proposed method with SOTA correspondence-based and correspondence-free registration methods.

##### Data generation

First, we randomly generate the source point cloud  $\mathcal{P}$  in the cube  $[-100, 100]^3$ . The source point cloud is transformed by a random rotation  $\mathbf{R}_{gt} \in \mathbb{SO}(3)$  and a random translation  $\mathbf{t}_{gt} \in [-100, 100]^3$  to generate the target point cloud  $\mathcal{Q}$ . Then, a portion of points in the target point cloud is replaced by arbitrarily generated points to simulate outliers. The outlier rate  $\eta$  is the ratio of these replaced points to all points. Besides, zero-mean Gaussian noise with standard deviation  $\sigma$  is added to the target point cloud. Notably, the inlier threshold in each synthetic data experiment is set according to the standard deviation of the noise.

##### Efficiency and accuracy experiments

This section presents three sets of experiments comparing the efficiency and accuracy of Ours with GORE, RANSAC, TEASER, FGR, GC-RANSAC, and TR-DE. Rotation errors, translation errors, and time costs are recorded for each experiment group. The first group focuses on experiments with a regular number of correspondences. We randomly generate  $N = \{1000, 2000, \dots, 5000\}$  correspondences with a noise level of  $\sigma = 0.5$  and an outlier rate of  $\eta = 0.5$ . The experiment is repeated 50 times for each setting, and the average results are depicted in Fig. 4.5. It is worth noting that results are not reported when the running time exceeds 1800 seconds. Among the deterministic methods, GORE and TEASER exhibit relatively high accuracy. However, their time costs increase significantly as the number of correspondences grows, with TEASER being the fastest in this regard. FGR, on the other hand, demonstrates occasional unsuccessful results but shows high efficiency. RANSAC and GC-RANSAC suffer from lower accuracy due to sampling uncertainty. Nevertheless, they exhibit relatively high efficiency at the regular outlier rate ( $\eta = 0.5$ ). In contrast, Ours outperforms all other methods in terms of both efficiency and accuracy. When  $N$  reaches 4000, Ours is approximately  $10^4$  times faster than GORE and TEASER. This may be explained by the reason that even after outlier rejection, a significant number of candidate inlier correspondences

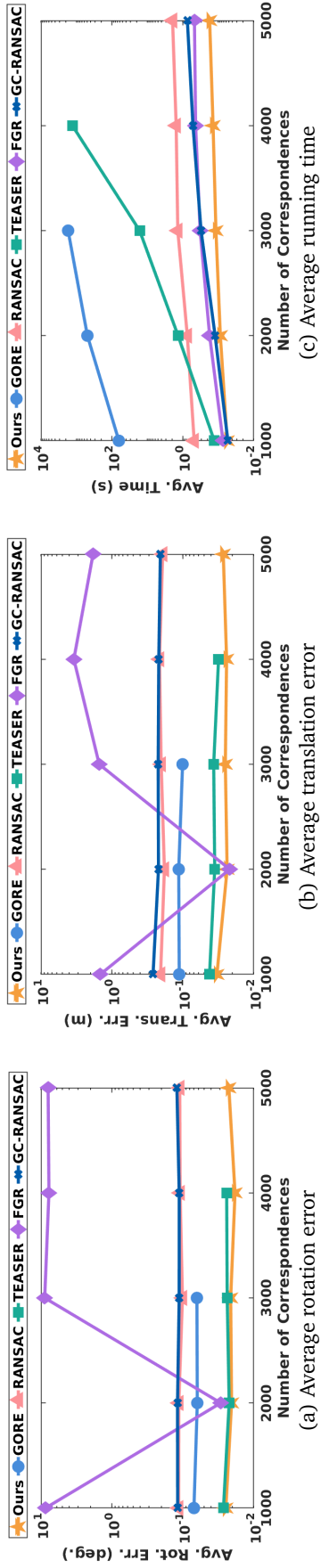


Figure 4.5: Controlled experiments with  $N = \{1000, 2000, \dots, 5000\}$ . The results include average rotation errors, average translation errors, and average running times.

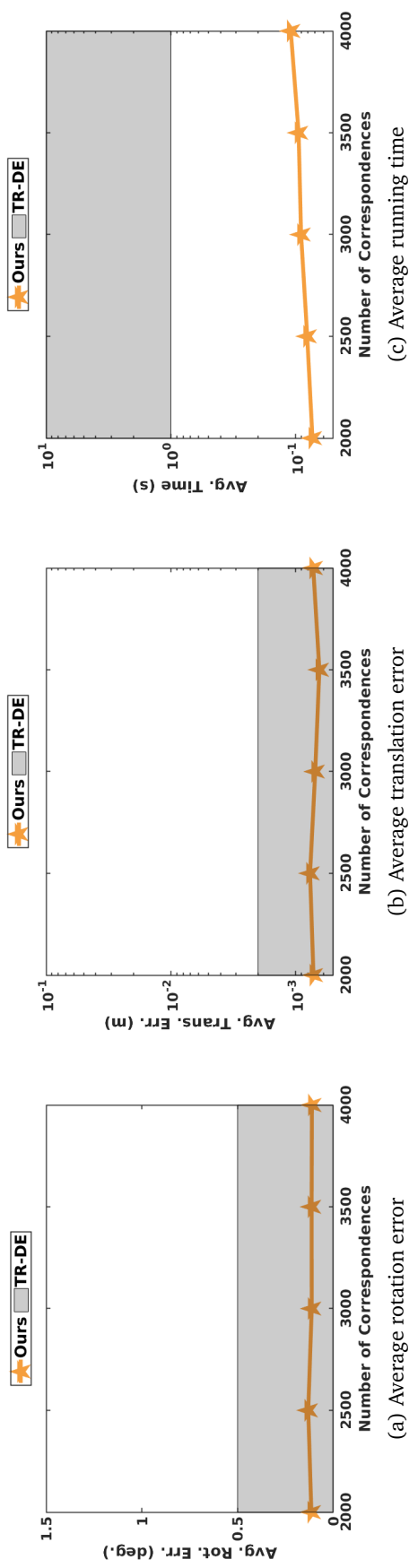


Figure 4.6: Controlled experiments with the same experimental conditions as TR-DE [Che+22c]. We use the gray rectangular region to approximately represent the results of TR-DE given in [Che+22c]. The results include average rotation errors, average translation errors, and average running times.

are still retained when dealing with a large number of correspondences. Consequently, the optimization process for GORE and TEASER becomes slower.

Since the code of TR-DE [Che+22c] is not released publicly, we set the same experimental conditions as TR-DE to compare the performance, which is the second group of experiments. Specifically, the source point cloud is randomly generated within the unit cube, and the experiment is conducted with  $N = \{2000, 2500, \dots, 4000\}$ ,  $\sigma = 0.005$ , and,  $\eta = 0.55$ . We also conduct 200 independent trials for each setting and record the average experiment results, as shown in Fig. 4.6. We use the gray rectangular region to approximately represent the results of TR-DE given in [Che+22c]. We can observe that Ours is about 10 times faster than TR-DE while keeping comparable accuracy.

To further investigate the potential efficiency advantages of Ours, we conduct the third group of experiments, specifically focusing on extremely high numbers of correspondences:  $N = \{10k, 20k, 50k, 100k, 200k, 500k\}$  (where  $k$  denotes one thousand). The remaining settings are consistent with those of the first group. Table 4.1 presents the average rotation errors, average translation errors, and average time costs of each method. The running time of GORE exceeds one hour starting from  $N = 10k$ , thus its results are not reported. Furthermore, TEASER demands a substantial amount of memory space, which renders it unable to operate efficiently under such extreme experimental conditions. As  $N$  increases to  $500k$ , RANSAC yields numerous unsatisfactory solutions and incurs a time cost of up to 315.7s. Additionally, GC-RANSAC fails to converge to the correct result after  $N$  reaches  $50k$  due to early termination. In comparison to FGR, Ours delivers more accurate rotation estimates but slightly less accurate translation estimates. However, experimental results indicate that the number of correspondences has a relatively minor impact on the efficiency of our method. For instance, when the number of correspondences increases from  $10k$  to  $500k$ , Ours is approximately 8 to 20 times faster than RANSAC and roughly 4 times faster than FGR. Overall, the proposed method exhibits superior efficiency while maintaining competitive accuracy compared to SOTA approaches.

### Robustness experiments

In this section, we conduct a group of controlled experiments to compare the robustness of Ours with GORE, RANSAC, TEASER, FGR, and GC-RANSAC. We randomly generate  $N = 2000$  correspondences with varying outlier rates ( $\eta = \{0.1, 0.2, \dots, 0.8\}$ ) and a noise level of  $\sigma = 0.5$ . The average rotation errors, average translation errors, and average time costs for each method are reported in Fig. 4.7. Results beyond a running time of 1800 seconds are not recorded in this group of experiments. Comparing the registration errors demonstrates that Ours, GORE, RANSAC, and TEASER are robust against up to 80% outlier rates. RANSAC has relatively higher registration errors than Ours, GORE, and TEASER. Moreover, the running time of RANSAC increases significantly with an increase in the outlier rate. In contrast, both GORE and TEASER display a significant decrease in running times as the outlier rate increases due to a corresponding reduction in the number of inliers. This indicates that, for GORE and TEASER, the time required for outlier removal is considerably smaller compared to the time spent on the optimization part. Consequently, they exhibit lower efficiency at regular outlier

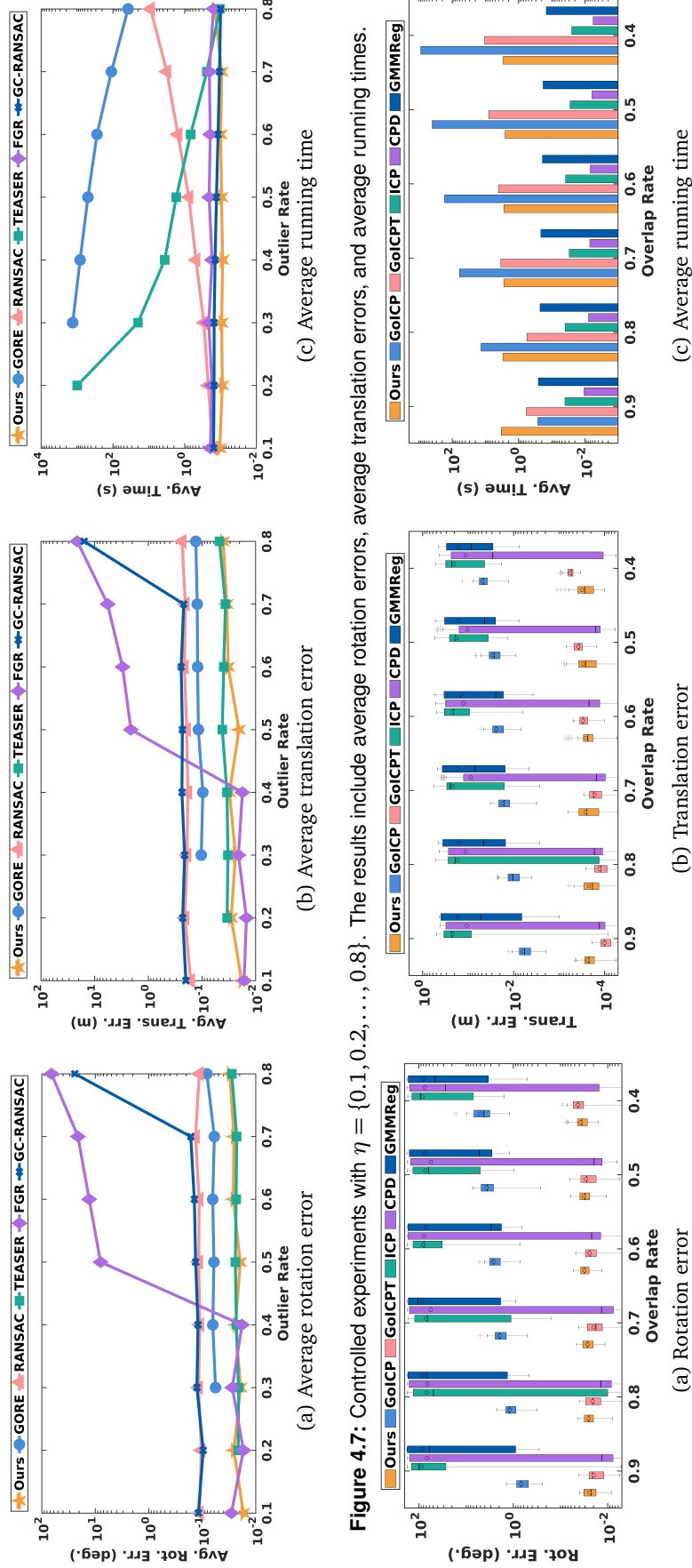


Figure 4.7: Controlled experiments with  $\eta = \{0.1, 0.2, \dots, 0.8\}$ . The results include average rotation errors, average translation errors, and average running times.

Figure 4.8: Controlled SPCR experiments with  $\rho = \{0.9, 0.8, \dots, 0.4\}$ . The results include rotation errors, translation errors, and average running times.

rates (e.g.,  $\eta \leq 0.5$ ). On the other hand, despite the high efficiency exhibited by both the deterministic FGR and the non-deterministic GC-RANSAC, they do not perform well when confronted with high outlier rates (e.g.,  $\eta \geq 0.7$ ). In contrast, Ours stands out as one of the fastest and most robust methods.

### Challenging SPCR experiments

In this section, we evaluate the performance of our extended simultaneous pose and correspondence registration (SPCR) method against GoICP, GoICPT, ICP, CPD, and GMMReg using the Bunny dataset [CL96]. The Bunny dataset consists of 35947 points and is pre-normalized to fit within the cube  $[-1, 1]^3$ , as required by GoICP [Yan+16]. Similar to [YSC20], we downsample the Bunny dataset to  $M = 100$  points, which serve as the source point cloud  $\mathcal{P}$ . To generate the target point cloud  $\mathcal{Q}$ , we apply a random rotation and translation to the source point cloud. Additionally, we randomly remove a certain proportion of points from  $\mathcal{Q}$  to simulate partial overlap between  $\mathcal{P}$  and  $\mathcal{Q}$ . The visualization results for a pair of synthetic data are shown in Fig. 4.1(d-1), where the bolded points represent the downsampled point clouds. Furthermore, we add zero-mean Gaussian noise with  $\sigma = 0.001$  to the source point cloud  $\mathcal{P}$ . The registration experiment is repeated 50 times for each overlap rate in  $\rho = \{0.9, 0.8, \dots, 0.4\}$ . Notably, the trimming fraction of GoICPT is set to be identical to the overlap rate.

The registration errors and average running times for each approach are presented in Fig. 4.8. Notably, the running times of building distance transform (DT) [Yan+16] for GoICP and GoICPT are not recorded and approximately take 23s on average. During repeated experiments, the local methods, ICP, CPD, and GMMReg, exhibit a tendency to converge to local optima, resulting in incorrect results. However, their efficiency remains a notable advantage. In contrast, the global methods GoICP and its variant GoICPT demonstrate greater robustness than these local methods. In particular, with a precisely tuned trimming ratio, GoICPT achieves remarkably high accuracy across all overlap rates. Nevertheless, these global methods suffer from relatively slow running times, which increase more rapidly than the proposed method. Consequently, when the overlap ratio is low (e.g.,  $\rho \leq 0.7$ ), Ours is faster than GoICP and GoICPT. As a global method, Ours also falls short in terms of efficiency compared to the local methods. However, Ours is more robust than local methods such as ICP, CPD, and GMMReg. Furthermore, as depicted in Fig. 4.1(d), Ours exhibits greater robustness than ICP and higher efficiency than GoICP on a randomly generated pair of Bunny data ( $\rho = 0.6$ ). These experiments illustrate the potential practicality of our proposed approach in addressing the challenging SPCR problem and its strength in terms of robustness and efficiency.

### 4.4.3 Real-World Data Experiments

In this section, to assess the performance of the proposed method on real-world data, we conduct experiments using the ETH dataset [TWS14] and the KITTI dataset [GLU12]. These datasets depict challenging outdoor LiDAR scenarios, with the former captured using terres-

**Table 4.2:** Detailed information about the ETH Dataset.

Scan pair	Number of points ( $10^6$ )	Number of keypoints	Number of correspondences	Outlier rate
Arch1	23.56-30.90	19007-12254	12617	98.45%
Arch2	30.90-29.45	12254-13286	11699	98.77%
Courtyard1	12.71-12.15	9634-12125	15325	86.55%
Courtyard2	12.15-16.75	12125-4081	8069	90.62%
Facade1	25.08-15.25	1586-2810	1901	97.16%
Facade2	15.25-15.79	2810-2215	2368	96.92%
Office1	10.73-10.69	1348-1277	1279	97.65%
Office2	10.69-10.75	1277-1486	1355	98.97%
Trees1	19.63-19.60	10883-10898	9543	99.41%
Trees2	20.39-20.48	12542-12522	11253	97.87%

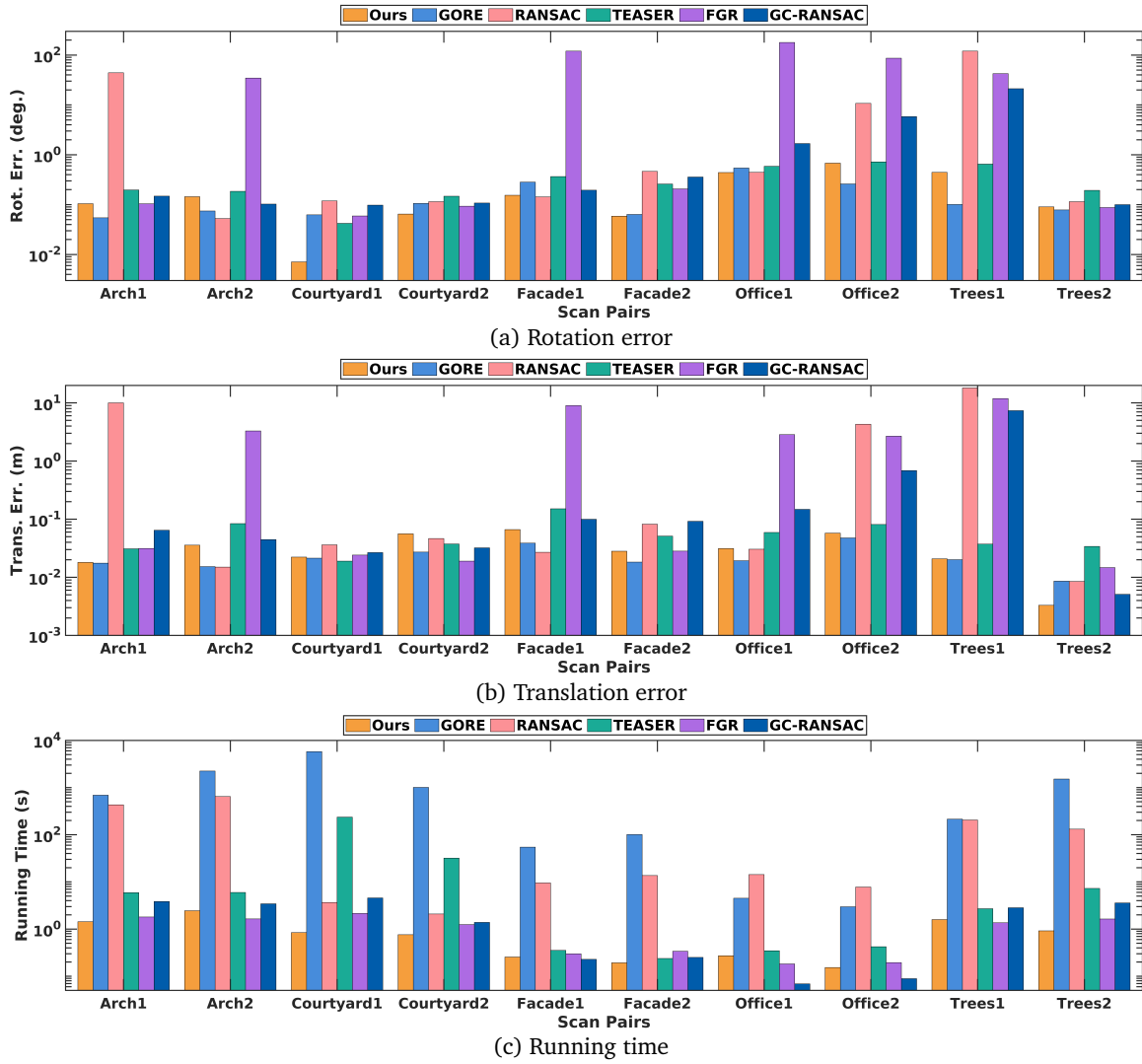
trial LiDAR and the latter collected from onboard LiDAR.

#### ETH dataset experiments

The ETH dataset [TWS14] is a challenging large-scale LiDAR dataset that encompasses five distinct scenarios: *Arch*, *Courtyard*, *Facade*, *Office*, and *Trees*. The average overlap rates of these scenarios are 30 – 40%, 40 – 70%, 60 – 70%, > 80%,  $\approx$  50%, respectively, as reported in [Li+23b]. To ensure the generality of the registration algorithm, we select two scan pairs from each scenario for our registration experiments. The ETH dataset provides ground truth information regarding the relative pose, enabling accurate evaluation. Similar to [Li22; Li+23b], we initially downsample the scans using the voxel grid algorithm [RC11]. Subsequently, we extract ISS [Zho09] keypoints and calculate FPFH [RBB09] descriptors for each keypoint. Through K-nearest neighbor search, we generate the set of putative correspondences  $\mathcal{K}$ . The downsampling resolution and the inlier threshold are both set to 0.1m. Detailed information about the ETH dataset, including the number of points, number of keypoints, number of correspondences, and outlier rate, can be found in Table 4.2. The outlier rate in the ETH dataset ranges from approximately 86% to 99%, with the number of correspondences varying from around 1k to 15k. To evaluate the registration performance, we compare Ours, GORE, RANSAC, TEASER, FGR, and GC-RANSAC using a total of 10 scan pairs from the ETH dataset.

Fig. 4.9 reports the rotation error, translation error, and running time for each method evaluated on the ETH dataset. Ours, GORE, and TEASER achieve remarkable robustness over all five scenes, successfully registering all scan pairs. GORE exhibits better accuracy overall compared to Ours, although it is time-consuming. Nevertheless, the registration errors achieved by Ours are still acceptable for practical applications. While the overall accuracy of TEASER is lower than that of Ours and GORE, its time cost increases significantly when dealing with a large number of inliers. For instance, Ours is approximately 280 times faster than TEASER in aligning the scan pair Courtyard1 with an outlier rate of 86.55%,

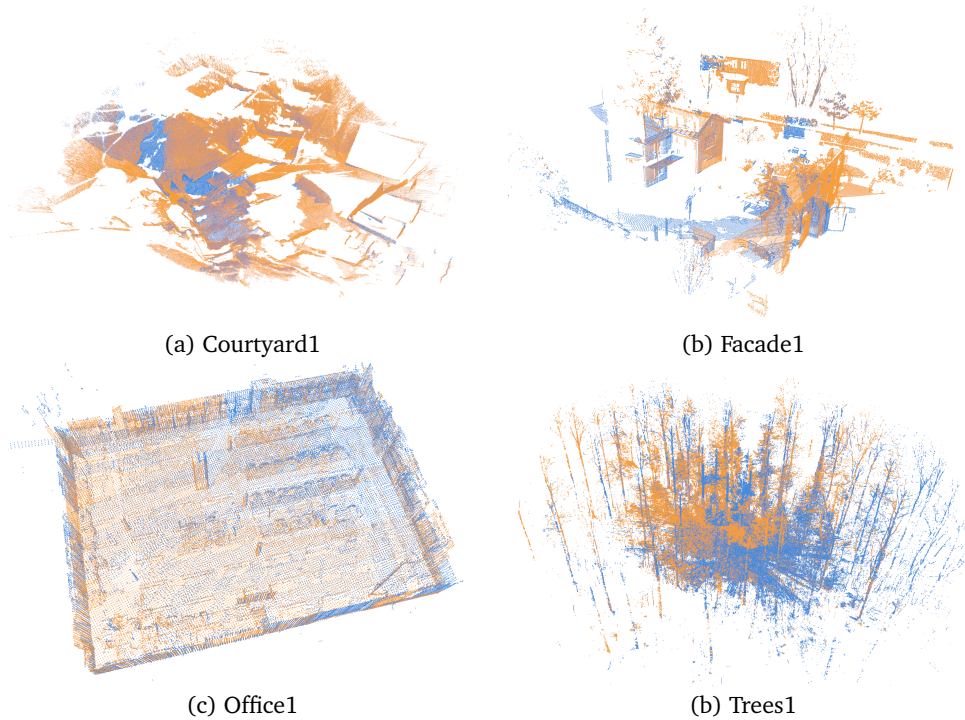




**Figure 4.9:** Experiment results on the ETH dataset [TWS14] with the PPFH [RBB09] descriptor. The results include rotation errors, translation errors, and running times.

and about 42 times faster than TEASER in aligning the scan pair Courtyard2 with an outlier rate of 90.62%. Another registration case for the scan pair Arch1 (with an outlier rate of 98.45%) is illustrated in Fig. 4.1(b), where Ours achieves the lowest translation error and is roughly 480 times faster than GORE and about 4 times faster than TEASER. The visualization results of the proposed method for the remaining four scenarios are provided in Fig. 4.10. Besides, FGR and GC-RANSAC demonstrate relatively high efficiency but exhibit instability when registering scan pairs with high outlier rates, such as Office2 and Trees1. RANSAC is not only time-consuming on the ETH dataset, but also prone to producing incorrect registration results. In summary, benefiting from the pose decoupling strategy based on residual projections, the proposed registration method is more efficient than the SOTA methods while maintaining robustness.





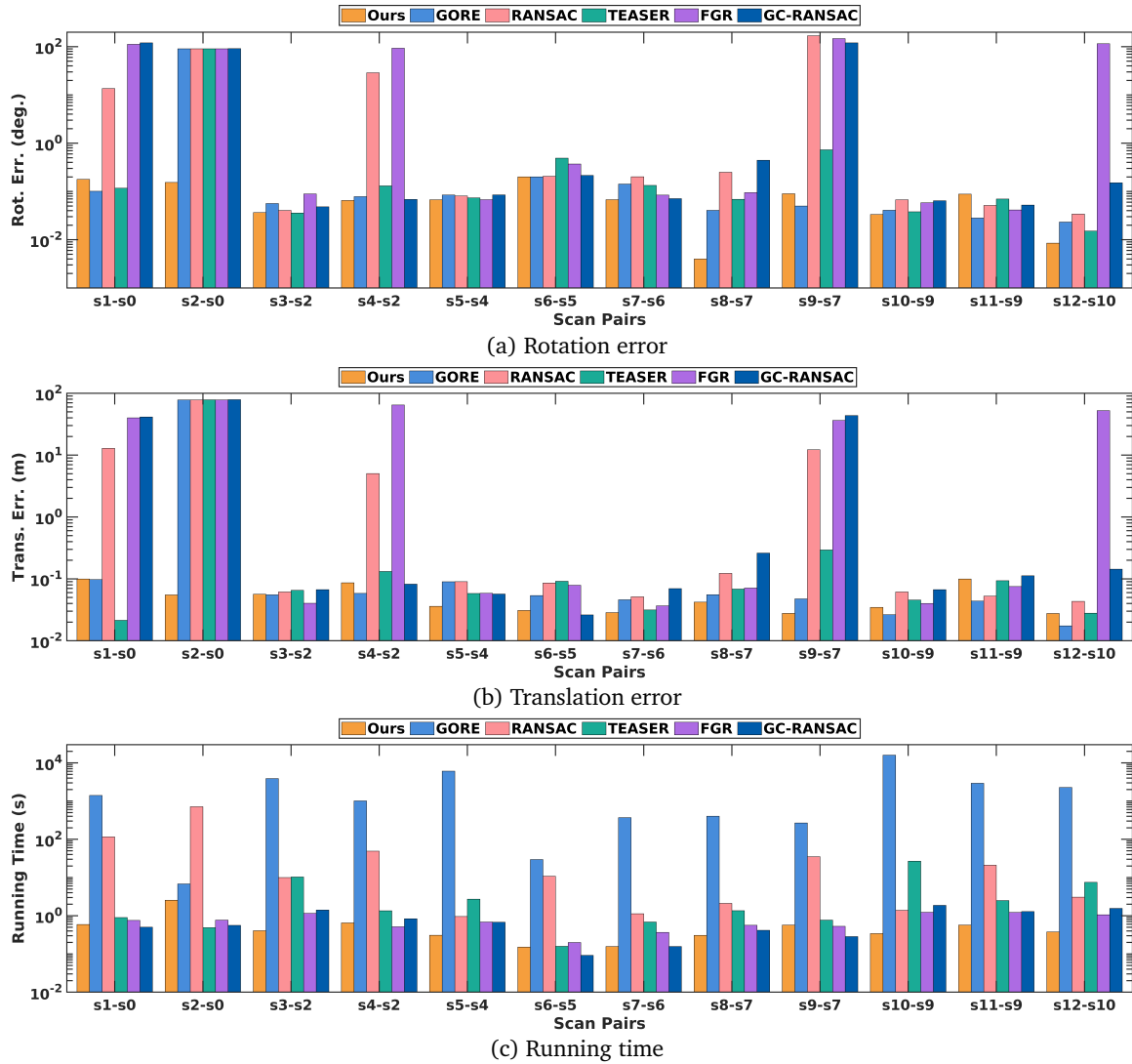
**Figure 4.10:** Registration results of the proposed method on the ETH dataset [TWS14], including four scan pairs: (a) Courtyard1, (b) Facade1, (c) Office1, and (d) Trees1. The aligned source point cloud is blue, and the target point cloud is yellow.

**Table 4.3:** Detailed information about the Bremen Dataset.

Scan pair	Number of points ( $10^6$ )	Number of keypoints	Number of correspondences	Outlier rate
s1-s0	16.16-15.90	30328-30290	5001	98.54%
s2-s0	15.25-15.90	39368-30290	6303	99.64%
s3-s2	15.03-15.25	43856-39368	8194	95.08%
s4-s2	18.05-15.25	26581-39368	5393	97.59%
s5-s4	18.76-18.05	20023-26581	4768	91.19%
s6-s5	20.33-18.76	9423-20023	1840	97.34%
s7-s6	18.47-20.33	16608-9423	2554	93.46%
s8-s7	15.85-18.47	19599-16608	3934	94.20%
s9-s7	16.29-18.47	32281-16608	4291	97.48%
s10-s9	15.18-16.29	36689-32281	8662	90.67%
s11-s9	14.61-16.29	37187-32281	7563	96.50%
s12-s10	15.76-15.18	36084-36689	8214	92.62%

#### Bremen dataset experiments and 3D Scene Reconstruction

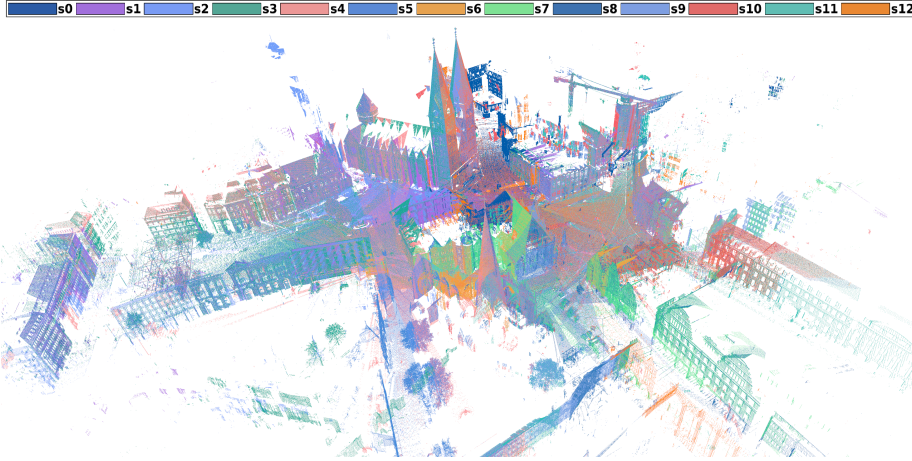
The Bremen dataset [BEN13] is a large-scale outdoor dataset with 13 LiDAR scans. We follow the same data preparation strategy outlined in the ETH dataset experiment to establish the initial correspondence set  $\mathcal{K}$ . The ground-truth pose for each scan is provided within the dataset. Since the proposed method is only for pair-wise registration, we construct 12



**Figure 4.11:** Experiment results on the Bremen dataset [BEN13] with the FPFH [RBB09] descriptor. The results include rotation errors, translation errors, and running times.

scan pairs to register all scans. Table 4.3 provides detailed information for each scan pair from the Bremen dataset, including the number of points, number of keypoints, number of correspondences, and outlier rate. The downsampling resolution for the Bremen dataset is set to 0.15m, which also determines the inlier threshold. With several thousand correspondences, the outlier rate ranges from approximately 90% to 99% for the Bremen dataset. We employ the proposed method (Ours), as well as GORE, RANSAC, TEASER, FGR, and GC-RANSAC, to register these scan pairs.

The rotation error, translation error, and running time of each method for each scan pair are shown in Fig. 4.11. Notably, when dealing with the registration of the scan pair s2-s0, all compared methods, except for Ours, fail due to the exceptionally high outlier rate (99.64%). GORE and TEASER demonstrate successful alignment with relatively high accuracy for the remaining scan pairs. Despite this, GORE exhibits the highest time cost among all methods, even when the number of correspondences is small or the outlier rate is low, which is consistent with the findings from synthetic data experiments. For instance, in the case of the



**Figure 4.12:** Our 3D scene reconstruction results on the Bremen dataset [BEN13], where different scans are indicated by different colors. The pair-wise point cloud registration is conducted for all 12 scan pairs.

pair s10-s9, which only has a 90.67% outlier rate, GORE requires over 3 hours for alignment, while TEASER takes up to 26.84 seconds. In contrast, Ours achieves registration in a mere 0.342 seconds. Furthermore, Fig. 4.1(a) shows another registration case for the scan pair s8-s7, where Ours not only achieves better accuracy but also is approximately  $10^3$  times faster than GORE and about 4 times faster than TEASER. Overall, Ours shows higher efficiency compared to GORE and TEASER, which exhibit similar levels of robustness and accuracy as Ours.

On the other hand, non-deterministic RANSAC demonstrates unstable performance, occasionally generating unsatisfactory solutions with significant registration errors, as observed in pairs s1-s0, s2-s0, s4-s2, and s9-s7. Moreover, RANSAC is also time-consuming in these practical scenarios with high outlier rates and a large number of correspondences. FGR, while fast for all scan pairs, often converges to erroneous results. Although GC-RANSAC outperforms RANSAC in terms of stability and efficiency, it still struggles to register all scan pairs successfully. In contrast, Ours exhibits remarkable robustness, achieving a 100% registration success rate on the Bremen dataset. The 3D scene reconstruction results on the Bremen dataset are given in Fig. 4.12.

#### KITTI dataset experiments

Following the data preparation strategy in [CPK19; Bai+21; Che+22c], we evaluate the performance of the proposed method on the KITTI dataset [GLU12]. The initial correspondences are generated using the learning-based descriptor FCGF [CPK19], and the inlier threshold is set to 0.6m. For successful registration, we set the thresholds for rotation error ( $E_R$ ) and translation error ( $E_t$ ) to  $5^\circ$  and 0.6m, respectively. In addition to comparing the performance of Ours against traditional methods such as RANSAC, TEASER, FGR, GC-RANSAC, and TR-DE, we also compare it with learning-based methods, including DGR [CDK20], DHVR [Lee+21], and PointDSC [Bai+21]. Notably, the learning-based descriptor FCGF outperforms traditional descriptors, resulting in a relatively low outlier rate for FCGF-based correspondences

**Table 4.4:** Experiment results on the KITTI dataset[GLU12] with FCGF [CPK19] descriptors. Bolded and underlined fonts indicate the first two best values.

Method	SR(%)	$E_R(^{\circ})$	$E_t$ (cm)	F1(%)	Time(s)
<i>i) Traditional</i>					
RANSAC[FB81]	96.40	0.36	21.12	84.77	2.56
TEASER[YSC20]	95.50	<u>0.33</u>	22.38	85.77	31.5
FGR[ZPK16]	96.94	0.34	<u>19.69</u>	85.80	0.99
GC-RANSAC[BM21]	97.48	<u>0.32</u>	20.68	85.42	1.16
TR-DE[Che+22c]	<u>98.20</u>	0.38	<b>18.00</b>	<u>85.99</u>	3.01
<i>ii) Deep learned</i>					
DGR[CDK20]	95.14	0.43	23.28	73.60	0.86
DHVR[Lee+21]	<b>99.10</b>	<b>0.29</b>	19.80	-	0.83
PointDSC[Bai+21]	97.84	0.33	20.99	85.29	<b>0.31</b>
Ours	<u>98.20</u>	<u>0.32</u>	20.05	<b>86.40</b>	<u>0.62</u>

(approximately 58.7% on average). Consequently, GORE is significantly slow on the KITTI dataset, so we do not report its results.

As shown in Table 4.4, all methods achieve a success rate exceeding 95% owing to the low outlier rate of FCGF-based correspondences. Among these methods, Ours attains the second-best success rate of 98.20%, slightly lower than DHVR with 99.10%. Although Ours is not the most efficient method, it ranks second in terms of efficiency among all approaches. For instance, Ours is approximately 5 times faster than the SOTA BnB-based TR-DE, about 4 times faster than the non-deterministic RANSAC, and approximately 50 times faster than the deterministic TEASER. It is worth mentioning that the most efficient method is the learning-based registration method, PointDSC. However, learning-based methods often require additional training procedures and may perform well only on the datasets they were trained on. Additionally, Ours exhibits the second-best rotation accuracy and the best  $F1$ -score. In Fig. 4.1(c), an example of registering a selected pair from the KITTI dataset is provided, where Ours has better accuracy and efficiency than FGR and GC-RANSAC. In general, when compared to SOTA methods, including learning-based methods, Ours showcases competitive performance in efficiency and robustness. This underscores the effectiveness of both the proposed pose decoupling strategy and the BnB-based search method.

### Orthogonality evaluation and verification

As the proposed method independently searches for the three rotation vectors without considering the orthogonality constraint, an evaluation of the orthogonality of the estimated coarse results is conducted on the KITTI dataset. Following the definition of a rotation matrix, the inner product for each pair of rotation vectors and the determinant of the coarse rotation matrix are computed, and the results are recorded in Table 4.5. Notably, the calculation of the mean and standard deviation of the metrics includes both successful and unsuccessful registration cases on the KITTI dataset. The findings indicate that, in most cases, the inner

**Table 4.5:** Orthogonality and determinant evaluation results on the KITTI dataset. The evaluation metrics comprise  $O_{XY}$ , defined as the inner product of  $\mathbf{r}_X^*$  and  $\mathbf{r}_Y^*$ , alongside  $O_{XZ}$ ,  $O_{YZ}$ , and the determinant  $Det$  of  $[\mathbf{r}_X^*, \mathbf{r}_Y^*, \mathbf{r}_Z^*]^T$ . The mean and standard deviation of these metrics are given.

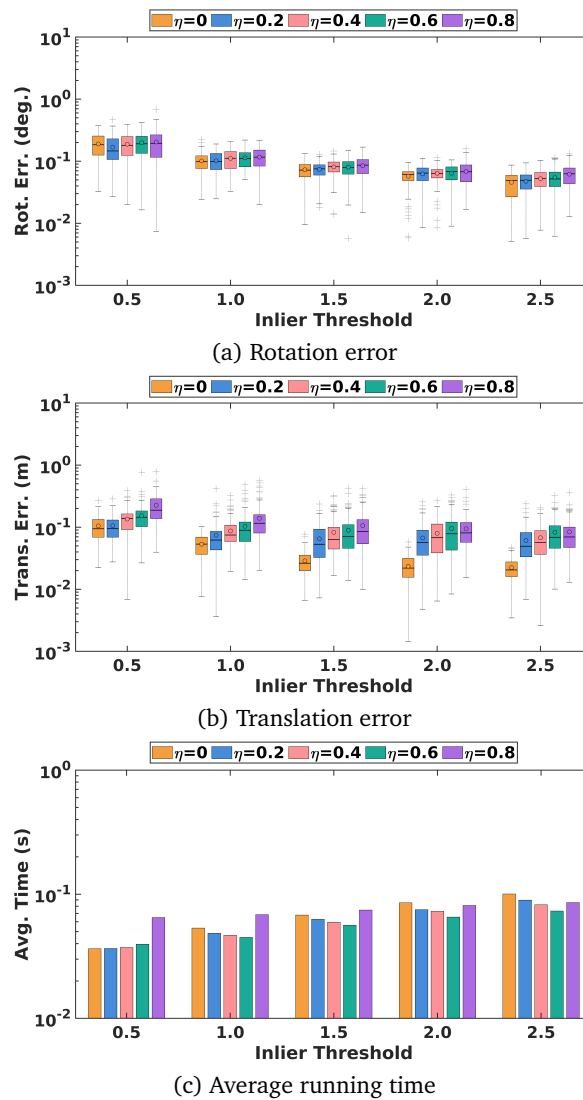
Metric	$O_{XY}$	$O_{XZ}$	$O_{YZ}$	$Det$
Mean	-0.0044	0.0031	-0.0019	0.9819
Standard deviation	0.0791	0.0990	0.1427	0.0763

product of any two rotation vectors is very close to 0, and the determinant of the coarse rotation matrix is likewise close to 1. Despite our method not explicitly considering orthogonality, its exceptional robustness (thanks to BnB) results in the coarse rotation matrix obtained from solving the sub-problems closely approximating the ground-truth rotation matrix in the majority of cases. This also indicates that the final consensus set obtained is very close to the true inlier set. As a consequence, the registration results after SVD refinement on the final consensus set are satisfactory (the registration success rate is 98.2%).

Additionally, thresholds of 0.3 for the inner product and 0.7 for the determinant are established to verify if the estimated solution of the decomposed sub-problems is valid. In the KITTI dataset (sequence 8 to 10), consisting of a total of 555 pairs of point clouds, the success rate of our method is 98.2%, implying 10 unsuccessful registration cases. Nonetheless, it is noteworthy that the estimated coarse matrices from 6 pairwise registration cases are verified as invalid rotation matrices. In this case, the SVD will also only yield erroneous results. Therefore, we can partially verify (6/10) if the solution is valid by checking the orthogonality of the coarse rotation vectors and the determinant of the coarse rotation matrix and thereby assess the quality of the estimations.

#### 4.4.4 Sensitivity Analysis

In this section, synthetic data experiments are conducted to theoretically assess the sensitivity of the proposed method to the inlier threshold. The inlier threshold is systematically adjusted to  $\lambda = \{1, 2, 3, 4, 5\}\sigma$ , while varying the outlier rate to  $\eta = \{0, 0.2, 0.4, 0.6, 0.8\}$ . We maintain  $N = 2000$  and a noise level of  $\sigma = 0.5$  throughout the experiments. Each configuration is replicated 100 times, and the results are presented in Fig. 4.13. The findings reveal that the inlier threshold  $\lambda$  only has a limited impact on registration accuracy and efficiency. Even with  $\lambda = 5\sigma$ , both the rotation error and translation error remain below  $1^\circ$  and 1m, respectively. However, increasing the inlier threshold leads to a slight increase in running time, potentially attributed to the larger inlier threshold resulting in a more relaxed upper bound, subsequently increasing the number of iterations.



**Figure 4.13:** Controlled experiments with different inlier thresholds  $\lambda = \{1, 2, 3, 4, 5\}\sigma$  and outlier rates  $\eta = \{0, 0.2, 0.4, 0.6, 0.8\}$ . The results include rotation errors, translation errors, and average running times.

## 4.5 Summary

This chapter presents an efficient, robust, and deterministic point cloud registration method, leveraging a novel pose decoupling strategy. By utilizing  $L_\infty$  residual projections, we successfully decouple the initial registration problem into three sub-problems, resulting in improved efficiency. Furthermore, we introduce a step-wise search strategy based on branch and bound for these sub-problems. Specifically, we define the inlier set maximization objective function and derive the novel upper bound based on the interval stabbing technique. Benefit from interval stabbing, we can additionally reduce the dimensionality of the branching space, thus accelerating the BnB search. Interestingly, thanks to its significant robustness, our proposed method can be extended to solve the challenging SPCR problem by introducing the interval merging technique. Extensive experiments conducted on both synthetic and real-world datasets demonstrate the competitive performance of our proposed method in terms of effi-

---

ciency and robustness when compared to SOTA approaches.





# 5

## Efficient and Robust LiDAR Registration with Gravity Prior

*This chapter aims to address the problem of robust point cloud registration with gravity prior, which is quite common in practice. The gravity directions are typically provided by inertial measurement units (IMUs) and can reduce the rotation to 1 degree of freedom (DOF). Commonly, existing outlier-robust global registration methods exhibit low efficiency. To overcome this limitation, this chapter introduces two novel registration algorithms based on transformation decoupling that allow searching for the solution within the lower-dimensional space, thereby significantly improving efficiency. The first algorithm utilizes known gravity directions to decompose the original 4-DOF problem into two sub-problems with 3-DOF and 1-DOF. Then, a deterministic global two-stage solver is presented to address the decoupled sub-problems sequentially. By further exploring the geometric properties, the second method employs the screw theory to decompose the raw 4-DOF problem into three sub-problems with 1-DOF, 2-DOF, and 1-DOF, respectively. A deterministic global three-stage solver is then developed. Extensive experiments demonstrate the superior performance of the two proposed methods in terms of efficiency and robustness.*

### 5.1 Background

Given 3D source and target point clouds, rigid point cloud registration is estimating the best transformation in  $\mathbb{SE}(3)$  that aligns two point clouds. It is also known as surface matching, an essential problem of computer vision and robotics. Point cloud registration has extensive applications in 3D reconstruction [BL95; Guo+20], pose estimation [Li+23d], object recognition [Guo+14; Xia+21], and robot localization [Li+23b; Zhu+22], etc.

Currently, researchers leverage prior information and reasonable assumptions to assist point cloud registration, such as planar motion, axis-fixed rotation, prior known gravity directions, etc [Cai+19; Don+20; Lim+22; Jia+22]. These pieces of information serve to reduce the dimensionality of the registration problem, particularly in terms of the param-

eters to be optimized, thereby enhancing algorithm efficiency. Among these prior pieces of information, the *gravity direction* (a.k.a. *vertical direction*) is extensively studied and serves as a shared direction to facilitate relative pose estimation [Din+20; Din+21; Liu+21], absolute pose estimation [HK17; Lec+19; LCK20a; LCK23], SLAM [Svä+16; Örn+22; KVP22], and panoramic stitching [DBK21; Bar+21]. With the aid of gravity direction, the relative rotation is reduced to 1 degree of freedom (DOF). Thus, the 6-DOF transformation in point cloud registration is reduced to only 4-DOF. In practice, modern autonomous systems, e.g., self-driving systems [Xu+22], commonly are equipped with inertial measurement units (IMUs), which can provide high precision gravity directions. On the other hand, the angular accuracy of the gravity direction is about  $0.5^\circ$  even for low-cost IMUs and less than  $0.02^\circ$  for high-accuracy IMUs [Din+20; Liu+21]. Alternatively, gravity directions can be obtained by vanishing point detection techniques from some structural scenarios [ADV03; LCK20b; Li+23a]. In this chapter, we focus on the general 4-DOF point cloud registration problem in which gravity directions are prior known. In particular, the registration for terrestrial LiDAR scans is a typical 4-DOF problem since the built-in tilt compensator keeps the rotation axis fixed [Cai+19; Don+20].

A classic paradigm for point cloud registration is estimating the transformation from a set of candidate correspondences generated by feature matching techniques and is also known as correspondence-based registration [BC17; Li22]. Nonetheless, due to the limited performance of current 3D matching methods and the challenges posed by partial overlap, noisy data, and structural duplication within real-world point clouds, the putative correspondences commonly contain a substantial number of outliers (often  $> 95\%$ ) [BC17; Li+23b]. To achieve robust registration, several common paradigms have been developed, of which the *consensus maximization* paradigm employed in our work is inherently robust to outliers and noise without smoothing or trimming to change the objective function [Li09; Cam+18]. Furthermore, consensus maximization has also been successfully adopted in many applications of robust fitting and estimation [CS17; Wen+19; TZN20; Liu+22].

The most popular heuristic used for solving the consensus maximization problem of point cloud registration is random sample consensus (RANSAC) [FB81]. RANSAC is an iterative hypothesize-and-test method that terminates after finding a consensus inlier set with sufficient correspondences or after achieving the total number of permitted iterations. However, RANSAC is non-deterministic and only generates a correct solution with a certain probability [Le+19]. Besides, the running time of RANSAC grows exponentially with the outlier rate [BC17]. To achieve highly outlier-robust registration, an increasing number of globally optimal and deterministic algorithms have been proposed. Most of the global methods [OKO08; HK09; BSP12; Yan+16; Bus+16; CP16; Cai+19] use branch-and-bound (BnB) to systematically search the entire solution domain. BnB-based methods can also solve the challenging problem of simultaneous pose and correspondence registration (SPCR) [Yan+16; Bus+16; CP16; Liu+18b; Cam+18; Liu+18a; Wan+21a], which is another paradigm for point cloud registration. BnB-based methods have two advantages: 1) they can find the globally optimal solution with theoretical guarantees, and 2) they are independent of initialization. However, their convergence speed is exponential to the dimensionality of the

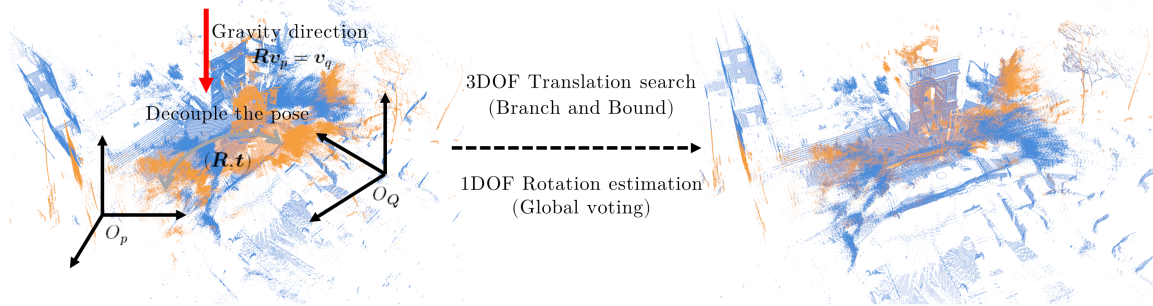
solution domain in the worst-case. Consequently, global methods are best appropriate for problems with small size and/or low dimensionality.

A popular strategy to improve the efficiency is *transformation decoupling* [BC17; Liu+18b; Str+17; Che+22c], by which the registration problem can be addressed within lower-dimensional parameter spaces. However, most existing methods address the general 6-DOF registration problem, with relatively fewer studies concentrating on the 4-DOF registration problem with gravity prior. Existing methods commonly decouple the original 6-DOF problem into two separate 3-DOF sub-problems [BC17; Str+17; Liu+18b; Li+18; Che+22c]. For instance, two special invariants are employed to decouple the rotation and translation, i.e., the rotation invariant features (RIFs) [Liu+18b; Wan+21a] and translation invariant measurements (TIMs) [Jia+22; YSC20]. Therefore, the algorithm efficiency still has potential room for improvement by considering the dimensionality of the registration problem.

## 5.2 Transformation Decoupling by Gravity Direction

This section proposes a novel, efficient, and deterministic registration method. The key idea is to decompose the joint 4-DOF pose into two sequential sub-problems with the aid of prior known gravity directions, i.e., 1) 3-DOF translation search, and 2) 1-DOF rotation estimation, as shown in Fig. 5.1. This decomposition mechanism reduces the search domain to 3-DOF, thus increasing efficiency. Specifically, a novel BnB-based method for the 3-DOF consensus maximization sub-problem is proposed, and then the corresponding lower and upper bound functions are derived. Next, an efficient global voting method for the 1-DOF rotation estimation sub-problem is proposed. Different from existing strategies based on RIFs [Liu+18b; Wan+21a] and TIMs [Jia+22; YSC20], which typically require at least two pairs of correspondences, our decoupling scheme is valid for a single pair of correspondences. In addition, since our target problem (4-DOF) is different from theirs (6-DOF), the core geometrical principles and operations in our work are also quite different. For example, the solution domain for rotation in the 6-DOF problem is  $\mathbb{SO}(3)$ , while in ours, it is  $[-\pi, \pi]$ . Furthermore, inspired by [YSC20], the SPCR problem is reformulated as a correspondence-based problem by assuming all-to-all correspondences and revising the objective and bound functions. Specifically, we assume that each point in the source point cloud is associated with all points of the target point cloud, and the impending problems have a large-scale set of candidate correspondences and extremely high outlier rates. However, the proposed method can also robustly solve this challenging SPCR problem.

Although both sub-problems are globally optimized in this study, the solution is not necessarily globally optimal for the raw 4-DOF point cloud registration problem but is deterministic. Notably, this is an inherent drawback common to the other decomposition-based approaches [Liu+18b; YSC20; Wan+21a; Che+22c]. The deterministic convergence of the proposed method is also essential for safety-critical applications like autonomous driving since the solutions given by the proposed method are guaranteed to be repeatable. In contrast, the results of heuristic methods such as RANSAC are probabilistic and non-deterministic.



**Figure 5.1:** Principle of our decoupled 4-DOF point cloud registration method. Specifically, the joint pose can be decoupled into the translation and rotation when the gravity directions are known. Our method comprises two procedures: 1) a 3-DOF translation search by BnB and 2) a 1-DOF rotation estimation by global voting. Since the search space of BnB is only 3-DOF, our method is able to significantly reduce the running time compared to the joint 4-DOF BnB-based method [Cai+19]. The illustration shows one pair of terrestrial LiDAR scans from *Arch* in the ETH dataset. After applying the fast match pruning (FMP) [Cai+19] to the candidate correspondences set, the running times of our method and the joint 4-DOF BnB-based method are 2.6ms and 635ms, respectively. Please note that the gravity directions used here are assumed as  $[0, 0, -1]^T$ .

The main contributions of this section can be summarized as follows:

- This section proposes a novel, simple, yet efficient decoupling scheme for the 4-DOF correspondence-based point cloud registration problem. Specifically, the 4-DOF problem is decoupled into two sub-problems with gravity prior. This strategy improves the convergence speed significantly by reducing the dimensionality of the solution domain.
- This section proposes a deterministic BnB-based optimization algorithm for the 3-DOF translation search sub-problem. Specifically, a new consensus maximization objective function is constructed with an angle-based criterion, and the new upper and lower bound functions are derived from the geometrical conditions. Furthermore, an efficient global voting algorithm is proposed for the 1-DOF rotation estimation sub-problem.
- Extensive experiments on several synthetic and real-world datasets demonstrate that the proposed method is not only more robust to outliers and noise than several existing heuristic methods, but also almost an order of magnitude faster than the existing joint 4-DOF BnB-based method. The proposed method strikes a balance between efficiency and robustness, promoting the BnB-based methods one step further in practical applications.
- This section extends the proposed method to address the challenging SPCR problem, employing the all-to-all correspondence assumption and accordingly developing a new objective. Experimental results show that the proposed method is more robust and accurate than several benchmark methods.

The rest of this section is organized as follows: Section 5.2.1 illustrates the problem formulation of the 4-DOF point cloud registration. Section 5.2.2 demonstrates the geometrical principle and details the proposed registration method.

### 5.2.1 Problem Formulation

#### Inlier Set Maximization

The source and target point clouds are defined as  $\mathcal{P}$  and  $\mathcal{Q}$ , respectively. In the correspondence-based registration setup, a set of candidate correspondences  $\mathcal{K} = \{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^N$  is extracted by matching points from  $\mathcal{P}$  and  $\mathcal{Q}$ , where  $\mathbf{p}_i, \mathbf{q}_i \in \mathbb{R}^3$  represent the point coordinates. The correspondences in  $\mathcal{K}$  tend to be contaminated by outliers and noise. The common 6-DOF registration problem is intended to estimate the rigid transformation  $T \in \mathbb{SE}(3)$ , including rotation  $\mathbf{R} \in \mathbb{SO}(3)$  and translation  $\mathbf{t} \in \mathbb{R}^3$ , so that the transformed point cloud  $T(\mathcal{P})$  is as close as possible to the point cloud  $\mathcal{Q}$ . In past studies, different forms of objective functions are employed, such as the  $L_2$  residual [Yan+16], Gaussian mixture model [CP16], and consensus maximization [Cai+19; Wan+21a; Che+22c]. In these objective functions, *consensus maximization* (a.k.a. *inlier set maximization*) is not only robust but also easy to compute. It is only necessary to compare the residual with the predefined threshold and count the number of inliers. Specifically, the optimization problem to be solved is:

$$T^* = \arg \max_{T \in \mathbb{SE}(3)} \mathcal{O}(T(\mathcal{P}), \mathcal{Q}), \quad (5.1)$$

where  $\mathcal{O}$  is the objective function for counting the number of inliers. Moreover, in the configurations of consensus maximization, different alignment criteria are applied, involving distance-based metrics [Bus+16; Liu+18a; Cai+19] and angle-based metrics [Cam+18; BWG19; Wan+21a], which then generate different problem formulations. In this section, the angle-based criterion is adopted to measure the alignment of point clouds. Specifically, assuming a candidate correspondence pair  $\mathbf{k}_i = (\mathbf{p}_i, \mathbf{q}_i)$  is an inlier, we obtain:

$$\mathbf{R}(\mathbf{p}_i + \mathbf{t}) = \mathbf{q}_i. \quad (5.2)$$

Then, given an arbitrary unit vector  $\mathbf{v} \in \mathbb{R}^3$ , the angle between the direction vector represented by each of the two points and this unit vector should be equal, i.e.,

$$\angle(\mathbf{R}(\mathbf{p}_i + \mathbf{t}), \mathbf{v}) = \angle(\mathbf{q}_i, \mathbf{v}). \quad (5.3)$$

Therefore, points  $\mathbf{p}_i$  and  $\mathbf{q}_i$  are considered as aligned (an inlier) only if  $|\angle(\mathbf{R}(\mathbf{p}_i + \mathbf{t}), \mathbf{v}) - \angle(\mathbf{q}_i, \mathbf{v})| \leq \zeta$ , where  $\zeta$  is the angle-based inlier threshold.

#### 4-DOF Registration

With the aid of known gravity directions provided by IMUs, the 6-DOF registration problem is reduced to 4-DOF. This is because the axes of coordinate systems can be aligned with the gravity direction so that the rotation is only 1-DOF. When the aligned axis is the Z-axis, the relative rotation is a pure yaw rotation. Now let us see the parameterization of rotation  $\mathbf{R}$ . The utilization of the geometric constraints introduced by the known gravity direction is considered. The gravity direction is a reference direction with unit-norm. Geometrically, the constraint of the gravity direction is given by:

$$\mathbf{R}\mathbf{v}_p = \mathbf{v}_q, \quad (5.4)$$

where  $\mathbf{v}_p$  and  $\mathbf{v}_q$  are unit gravity directions in the source and target point clouds  $\mathcal{P}$ ,  $\mathcal{Q}$ , respectively. The solution to this equation is:

$$\mathbf{R} = \mathbf{R}(\theta, \mathbf{v}_q) \cdot \mathbf{R}_{\mathbf{v}_p}^{\mathbf{v}_q}, \quad (5.5)$$

where  $\mathbf{R}_{\mathbf{v}_p}^{\mathbf{v}_q}$  is the rotation that aligns  $\mathbf{v}_p$  to  $\mathbf{v}_q$  with the minimum geodesic motion,  $\mathbf{R}(\theta, \mathbf{v}_q)$  is the rotation that rotates  $\theta \in [-\pi, \pi]$  about axis  $\mathbf{v}_q$ , and  $\theta$  is an unknown scalar (a 1-DOF problem). Please refer to [BC17] for more details regarding this solution. Eq. (5.5) indicates that the rotation is reduced from 3-DOF to 1-DOF. Therefore, given an arbitrary unit vector  $\mathbf{v} \in \mathbb{R}^3$ , the 4-DOF registration problem is given by:

$$\theta^*, \mathbf{t}^* = \arg \max_{\theta \in [-\pi, \pi], \mathbf{t} \in \mathbb{R}^3} \sum_{i=1}^N \mathbb{I}(|\angle(\mathbf{R}(\mathbf{p}_i + \mathbf{t}), \mathbf{v}) - \angle(\mathbf{q}_i, \mathbf{v})| \leq \zeta), \quad (5.6)$$

where  $\mathbb{I}$  is an indicator function that returns 1 if the input condition is true and 0 otherwise, and  $\zeta$  is the angle-based inlier threshold. The decoupling scheme will be introduced in the next section.

#### 4-DOF Transformation Decoupling by Gravity Direction

If consider the known gravity direction  $\mathbf{v}_q$  to be the arbitrary unit vector  $\mathbf{v}$  in Eq. (5.6), then the objective function is:

$$E(\mathbf{R}, \mathbf{t} | \mathcal{K}, \zeta) = \sum_{i=1}^N \mathbb{I}(|\angle(\mathbf{R}(\mathbf{p}_i + \mathbf{t}), \mathbf{v}_q) - \angle(\mathbf{q}_i, \mathbf{v}_q)| \leq \zeta), \quad (5.7)$$

where  $\mathbf{R}$  represents the 1-DOF rotation, and  $\mathcal{K}$  represents the set of candidate correspondences. According to Eq. (5.4), we have:

$$\begin{aligned} \angle(\mathbf{R}(\mathbf{p}_i + \mathbf{t}), \mathbf{v}_q) &= \angle(\mathbf{R}(\mathbf{p}_i + \mathbf{t}), \mathbf{R}\mathbf{v}_p) \\ &= \angle(\mathbf{p}_i + \mathbf{t}, \mathbf{v}_p). \end{aligned} \quad (5.8)$$

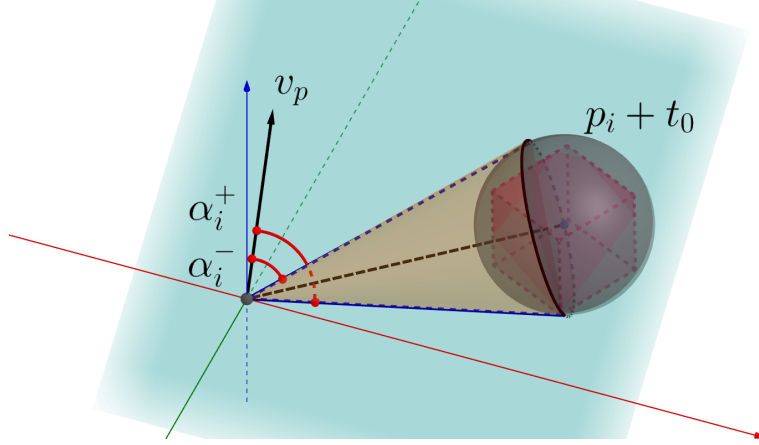
Eq. (5.8) results from the simple fact that when two 3D direction vectors are rotated together, the angle between the two vectors remains invariant. Since  $\mathbf{q}_i$  and  $\mathbf{v}_q$  are known, then we can define  $\gamma_i = \angle(\mathbf{q}_i, \mathbf{v}_q)$ . Thus the objective function (5.7) can be rewritten as:

$$E(\mathbf{t} | \mathcal{K}, \zeta) = \sum_{i=1}^N \mathbb{I}(|\angle(\mathbf{p}_i + \mathbf{t}, \mathbf{v}_p) - \gamma_i| \leq \zeta). \quad (5.9)$$

The objective function (5.9) is only related to the unknown translation  $\mathbf{t}$ . Regarding the parameterization of the 3-DOF translation,  $\mathbf{t}$  can be regarded as lying in the translation uncertainty cube  $\mathcal{C}_t$  with center  $\mathbf{t}_0$  and radius  $\rho_t$ , i.e.,  $\mathbf{t} \in [\mathbf{t}_0 - \rho_t, \mathbf{t}_0 + \rho_t]$ .

To sum up, the constraint of known gravity direction is utilized to decouple the 4-DOF consensus maximization problem into 3-DOF translation and 1-DOF rotation sub-problems.





**Figure 5.2:** Geometric principle of the proposed upper and lower bounds for BnB. The uncertainty cube  $\mathcal{C}_t^{P_i}$  is the red cube. The center of the cube is  $\mathbf{p}_i + \mathbf{t}_0$ . The light gray sphere is the relaxed circumsphere of the cube. The bound for the uncertain angle  $\alpha_i$  is  $[\alpha_i^-, \alpha_i^+]$ . The splitting policy is to divide from the center of each dimension of the translation domain, i.e., there will be  $2^3 = 8$  sub-branches after one split.

In the next section, an efficient two-stage solver for these sub-problems will be developed.

## 5.2.2 Decoupled 4-DOF Point Cloud Registration

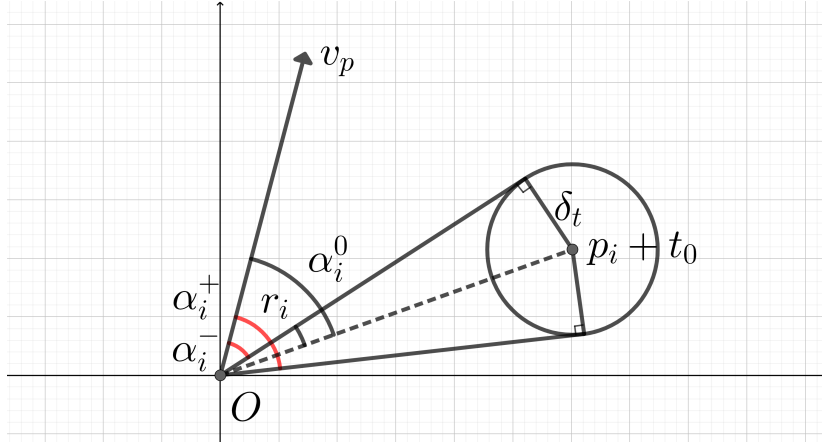
### Global Translation Search

BnB-based algorithms operate according to two principles: 1) branch: they recursively divide the search domain into smaller sub-branches, and 2) bound: they evaluate the lower and upper bounds on the optimal solution in each constrained sub-branch. The BnB-based algorithms then use these lower and upper bounds to prune the search domain and delete the sub-branch that can be proved not to contain the optimal solution. The BnB-based algorithms converge when the lower and upper bounds of the optimal solution are equal. Therefore, the key to BnB-based methods for solving the registration problem is to construct the corresponding upper and lower bounds.

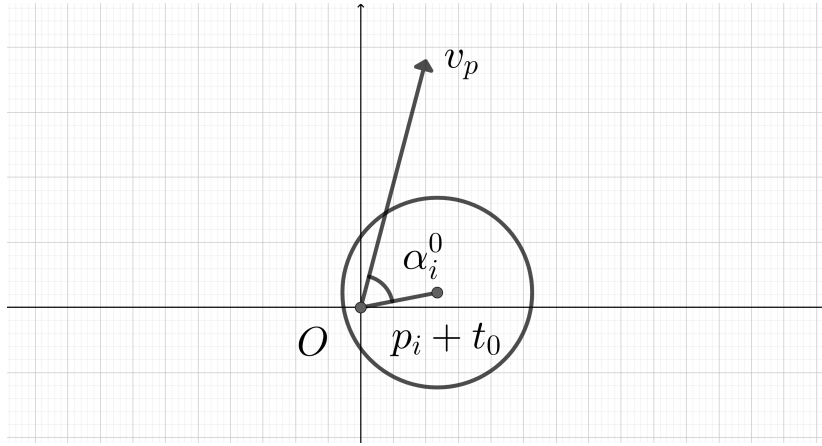
In order to bound the objective function (5.9) on an arbitrary branch, the bound for angle  $\angle(\mathbf{p}_i + \mathbf{t}, \mathbf{v}_p)$  needs to be found based on the uncertainty cube  $\mathcal{C}_t^{P_i}$ . First, the cuboidal region of  $\mathbf{p}_i + \mathbf{t}$  with radius  $\rho_t$  can be relaxed into a circumsphere region with radius  $\delta_t = \sqrt{3}\rho_t$ . Then, given this circumsphere region, the uncertain angle can be defined as  $\alpha_i = \angle(\mathbf{p}_i + \mathbf{t}, \mathbf{v}_p)$ , and the bound of  $\alpha_i$  can be derived geometrically, i.e.,  $\alpha_i^- \leq \alpha_i \leq \alpha_i^+$ . Specifically,  $\alpha_i^-$  and  $\alpha_i^+$  are the minimum and maximum angles between the tangents of the relaxed circumsphere region and the vector  $\mathbf{v}_p$ , respectively, as shown in Fig. 5.2. The tangents are constructed between the relaxed circumsphere region and the origin, with a conical shape, shown in orange in Fig. 5.2. Then  $\alpha_i^0$  is defined as the middle point of the interval  $[\alpha_i^-, \alpha_i^+]$ , and we can obtain:

$$\begin{aligned} \alpha_i^0 &= (\alpha_i^- + \alpha_i^+)/2 \\ &= \angle(\mathbf{p}_i + \mathbf{t}_0, \mathbf{v}_p). \end{aligned} \quad (5.10)$$

Moreover,  $r_i$  is defined as the radius of the interval  $[\alpha_i^-, \alpha_i^+]$ , such that  $r_i$  is the half conic



**Figure 5.3:** A 2D illustration of the upper bound in Case 1, i.e., the relaxed circumsphere and the origin do not intersect.



**Figure 5.4:** A 2D illustration of the upper bound in Case 2, i.e., the relaxed circumsphere and the origin intersect. The uncertain angle  $\alpha_i$  lies arbitrarily within the interval  $[-\pi, \pi]$ .

angle from the geometric perspective, as shown in Fig. 5.3. However, there is another case in which the relaxed circumsphere intersects with the origin, such that  $r_i$  is  $\pi$ , as shown in Fig. 5.4. This case means that the uncertain angle  $\alpha_i$  is arbitrary within the interval  $[-\pi, \pi]$ . Overall, there are two cases of  $r_i$ : either that the circumsphere and the origin do not intersect (Case 1), as shown in Fig. 5.3, or the circumsphere and the origin intersect (Case 2), as shown in Fig. 5.4, i.e.:

$$r_i = \begin{cases} \arcsin(\delta_t / \|\mathbf{p}_i + \mathbf{t}_0\|), & \text{if } \delta_t \leq \|\mathbf{p}_i + \mathbf{t}_0\| \\ \pi, & \text{otherwise.} \end{cases} \quad (5.11)$$

where  $\delta_t = \sqrt{3}\rho_t$ . Numerically, the minimum distance between  $\gamma_i$  and interval  $[\alpha_i^-, \alpha_i^+]$  is defined as  $l = \min |\gamma_i - [\alpha_i^-, \alpha_i^+]|$ , and we obtain:

$$|\alpha_i - \gamma_i| \geq l. \quad (5.12)$$

Eq. (5.12) is due to  $\alpha_i^- \leq \alpha_i \leq \alpha_i^+$ . When  $\gamma_i$  is not in the interval  $[\alpha_i^-, \alpha_i^+]$ ,  $l$  is equal to



the distance from the point  $\gamma_i$  to the middle point  $\alpha_i^0$  of the interval minus the radius of the interval  $r_i$ , i.e.:

$$l = |\alpha_i^0 - \gamma_i| - r_i \geq 0. \quad (5.13)$$

On the other hand, when  $\gamma_i$  is in the interval  $[\alpha_i^-, \alpha_i^+]$ , we obtain:

$$|\alpha_i^0 - \gamma_i| - r_i \leq 0 \leq l. \quad (5.14)$$

Therefore, according to Eq. (5.13) and Eq. (5.14), we obtain:

$$l \geq |\alpha_i^0 - \gamma_i| - r_i. \quad (5.15)$$

Then, combined with Eq. (5.12), we obtain:

$$|\alpha_i - \gamma_i| \geq |\alpha_i^0 - \gamma_i| - r_i. \quad (5.16)$$

Finally, the upper bound function on the sub-branch  $\mathbb{B} \subset \mathbb{R}^3$  is derived as follows:

$$\bar{E}(\mathbb{B}) = \sum_{i=1}^N \mathbb{I}(|\alpha_i^0 - \gamma_i| - r_i \leq \zeta). \quad (5.17)$$

In this upper bound function, instead of first calculating the circumsphere and then calculating the cone of the external tangent, we can directly calculate  $\alpha_i^0$  and  $r_i$ , which is given by Eq. (5.10) and Eq. (5.11). In addition, the lower bound function can be:

$$\underline{E}(\mathbb{B}) = \sum_{i=1}^N \mathbb{I}(|\alpha_i^0 - \gamma_i| \leq \zeta). \quad (5.18)$$

The following proposition and proof show the validity of Eq. (5.17) and Eq. (5.18).

**Proposition 5.1.** *For any translation sub-branch  $\mathbb{B} \subset \mathbb{R}^3$  with center  $t_0$  and radius  $\rho_t$ , the upper bound and lower bound of the objective function (5.9) can be chosen as  $\bar{E}(\mathbb{B})$  and  $\underline{E}(\mathbb{B})$  from Eq. (5.17) and Eq. (5.18), respectively.*

*Proof.* To prove the validity of the upper bound  $\bar{E}(\mathbb{B})$  and lower bound  $\underline{E}(\mathbb{B})$ , there are three issues that need to be addressed. Specifically, the maximum cardinality of the inlier set in any sub-branch  $\mathbb{B}$  should lie between the lower and upper bounds, i.e.,

$$\bar{E}(\mathbb{B}) \geq \max_{t \in \mathbb{B}} E(t | \mathcal{K}, \zeta) \geq \underline{E}(\mathbb{B}). \quad (5.19)$$

Additionally, the lower and upper bounds should be equal when sub-branch  $\mathbb{B}$  collapses to a single point. The following is the detailed proof procedure.

- Observe that  $\forall t \in \mathbb{B}$ , and according to Eq. (5.16), when  $(p_i, q_i)$  contributes 1 to the objective function (5.9), i.e.,

$$\mathbb{I}(|\alpha_i - \gamma_i| \leq \zeta) = 1, \quad (5.20)$$

$(p_i, q_i)$  must contribute 1 to Eq. (5.17), i.e.,

$$\mathbb{I}(|\alpha_i^0 - \gamma_i| - r_i \leq \zeta) = 1. \quad (5.21)$$

On the other hand, when  $(\mathbf{p}_i, \mathbf{q}_i)$  contributes 0 to the objective function (5.9), i.e.,

$$\mathbb{I}(|\alpha_i - \gamma_i| \leq \zeta) = 0, \quad (5.22)$$

$(\mathbf{p}_i, \mathbf{q}_i)$  probably contributes 1 or 0 to Eq. (5.17), i.e.,

$$\mathbb{I}(|\alpha_i^0 - \gamma_i| - r_i \leq \zeta) = 1 \text{ or } 0. \quad (5.23)$$

Thus, for all correspondences  $(\mathbf{p}_i, \mathbf{q}_i)$  in  $\mathcal{K}$ , we obtain:

$$\bar{E}(\mathbb{B}) = \sum_{i=1}^N \mathbb{I}(|\alpha_i^0 - \gamma_i| - r_i \leq \zeta) \geq \max_{\mathbf{t} \in \mathbb{B}} E(\mathbf{t} | \mathcal{K}, \zeta). \quad (5.24)$$

Therefore,  $\bar{E}(\mathbb{B})$  is the upper bound of the objective function  $E(\mathbf{t} | \mathcal{K}, \zeta)$ .

- The objective function  $E(\mathbf{t} | \mathcal{K}, \zeta)$  at an arbitrary point within sub-branch  $\mathbb{B}$  is less than the maximum of the objective function within sub-branch  $\mathbb{B}$ . We can easily pick the center of the sub-branch, i.e.,  $\mathbf{t}_0$ . Then, we obtain:

$$\underline{E}(\mathbb{B}) = \sum_{i=1}^N \mathbb{I}(|\alpha_i^0 - \gamma_i| \leq \zeta) \leq \max_{\mathbf{t} \in \mathbb{B}} E(\mathbf{t} | \mathcal{K}, \zeta). \quad (5.25)$$

Therefore,  $\underline{E}(\mathbb{B})$  is the lower bound of the objective function  $E(\mathbf{t} | \mathcal{K}, \zeta)$ .

- When the sub-branch  $\mathbb{B}$  collapses to a single point  $\mathbf{t}_s$ , and radius  $\rho_t = 0$ , we have  $r_i = 0$ , i.e.,

$$\lim_{\rho_t \rightarrow 0} (\bar{E}(\mathbb{B}) - \underline{E}(\mathbb{B})) = 0. \quad (5.26)$$

Therefore, the gap between the upper bound and lower bound is equal to zero, and the convergence of  $\bar{E}(\mathbb{B})$  and  $\underline{E}(\mathbb{B})$  is proved. □

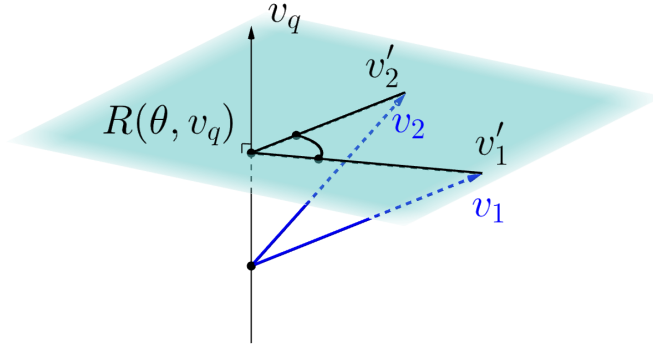
The BnB-based 3-DOF translation search algorithm is outlined in Algorithm 4 according to the proposed lower and upper bound in Proposition 5.1. The depth-first search strategy [MSS08] is employed, and the branch with the maximum upper bound is divided into eight sub-branches at each iteration, as shown in Line 6. The point of Line 7 is to delete the branch that has been split. Line 8 updates  $U$  to the maximum of the current upper bound of all branches (the current maximum), while Line 9 updates  $L$  to the maximum of the lower bound of all branches in all iterations (the historical maximum). The pruning process in Line 10 improves the convergence efficiency of the algorithm, which deletes the branches with upper bounds less than the historical maximum lower bound  $L$ . After the algorithm converges,

**Algorithm 4:** BnB for 3-DOF translation search (solution to problem (5.9))**Input:** Solution domain  $\mathbb{B}$ , Inlier threshold  $\zeta$ , Set of candidate correspondences

$$\mathcal{K} = \{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^N, \text{ and Gravity directions } \mathbf{v}_p, \mathbf{v}_q.$$

**Output:** Globally optimal solution  $\mathbf{t}^*$ .

- 1 Let  $\xi$  be the list of sub-branches, initialize  $\xi = \{\mathbb{B}\}$ ;
- 2 Define function  $\delta(\mathbb{B})$  returns the center point of sub-branch  $\mathbb{B}$ ;
- 3 Define function  $S(\mathbb{B})$  splits the sub-branch  $\mathbb{B}$ ;
- 4 Initialize  $L = 0$ , and  $U = N$ ;
- 5 **while**  $U - L > 0$  **do**
- 6     Select a sub-branch  $\mathbb{B}$  with the maximum upper bound from  $\xi$ , i.e.,  
 $\mathbb{B} = \arg \max \bar{E}(\mathbb{B}_k), \mathbb{B}_k \in \xi$ . Then split  $\mathbb{B}$  into eight sub-branches  
 $S(\mathbb{B}) = \{\mathbb{B}_1, \dots, \mathbb{B}_8\}$ ;
- 7     Delete  $\mathbb{B}$  from  $\xi$ , and add  $\{\mathbb{B}_1, \dots, \mathbb{B}_8\}$  to  $\xi$ ;
- 8     Update  $U = \max \bar{E}(\mathbb{B}_k), \mathbb{B}_k \in \xi$ ;
- 9     Update  $L = \max\{L, \underline{E}(\mathbb{B}_k)\}$  with  $\mathbb{B}_k \in \xi$ . Meantime, if  $\underline{E}(\mathbb{B}_k) > L$ , set  $\mathbf{t}^* = \delta(\mathbb{B}_k)$ ;
- 10    Delete  $\mathbb{B}_k$  from  $\xi$  with  $\bar{E}(\mathbb{B}_k) < L, \mathbb{B}_k \in \xi$ ;
- 11 **end**



**Figure 5.5:** Geometric intuition for angle  $\theta$ , which is the angle between the projections of two vectors  $\mathbf{v}_1 = \mathbf{R}_{\mathbf{v}_p}^{\mathbf{v}_q} \cdot (\mathbf{p}_i + \mathbf{t})$  and  $\mathbf{v}_2 = \mathbf{q}_i$ . The projections  $\mathbf{v}'_1$  and  $\mathbf{v}'_2$  are on the plane with  $\mathbf{v}_q$  as the normal.

the globally optimal solution for the translation can be obtained. In addition, the algorithm outputs the best-so-far solution in each iteration, as shown in Line 9, and it can still provide a best-guess solution when dealing with problems with limited running time, even if the upper and lower bounds are not equal.

### Global Rotation Estimation

In this study, the axis-angle form instead of the Euler angle form is used to represent the rotation according to the alignment of axes in the 4-DOF scenarios. Then Eq. (5.5) is substituted into Eq. (5.2):

$$\mathbf{R}(\theta, \mathbf{v}_q) \cdot \mathbf{R}_{\mathbf{v}_p}^{\mathbf{v}_q} \cdot (\mathbf{p}_i + \mathbf{t}) = \mathbf{q}_i, i = 1, \dots, N, \quad (5.27)$$

where

$$\mathbf{R}(\theta, \mathbf{v}_q) = \exp(\theta[\mathbf{v}_q]_{\times}) \quad (5.28)$$

$$= \mathbf{I} + \sin \theta[\mathbf{v}_q]_{\times} + (1 - \cos \theta)[\mathbf{v}_q]_{\times}^2, \quad (5.29)$$

and  $[\mathbf{v}_q]_{\times}$  returns the cross product matrix of  $\mathbf{v}_q$ . The solution to Eq. (5.27) is the rotation angle  $\theta$  such that the vector  $\mathbf{v}_1 = \mathbf{R}_{\mathbf{v}_p}^{\mathbf{v}_q} \cdot (\mathbf{p}_i + \mathbf{t})$  can rotate  $\theta$  around axis  $\mathbf{v}_q$  to the vector  $\mathbf{v}_2 = \mathbf{q}_i$ . Geometrically,  $\theta$  is the angle between the projections  $\mathbf{v}'_1$  and  $\mathbf{v}'_2$  of these two vectors on the plane with axis  $\mathbf{v}_q$  as the normal, i.e.,  $\theta = \angle(\mathbf{v}'_1, \mathbf{v}'_2)$ . The geometric intuition is shown in Fig. 5.5. Moreover, the rotation  $\mathbf{R}_{\mathbf{v}_p}^{\mathbf{v}_q}$  that aligns  $\mathbf{v}_p$  to  $\mathbf{v}_q$  with the minimum geodesic motion is:

$$\mathbf{R}_{\mathbf{v}_p}^{\mathbf{v}_q} = \exp(\rho[\mathbf{v}_m]_{\times}), \quad (5.30)$$

where  $\rho = \arccos(\mathbf{v}_p \cdot \mathbf{v}_q)$ , and  $\mathbf{v}_m = \frac{\mathbf{v}_p \times \mathbf{v}_q}{\|\mathbf{v}_p \times \mathbf{v}_q\|}$ .

The optimal translation  $\mathbf{t}^*$  is obtained from the last section. Therefore, a candidate rotation angle  $\theta_i$  for every single correspondence  $(\mathbf{p}_i, \mathbf{q}_i)$  can be calculated according to Eq. (5.27). Thus, the histogram voting [Che+09] can be directly leveraged in the interval  $[-\pi, \pi]$  to find the globally optimal solution. Specifically, the interval is equally divided into 360 segments, and then the candidate rotation angles  $\{\theta_i\}_{i=1}^N$  are distributed into these 360 segments. The midpoint of the segment with the maximum number of  $\theta_i$  is approximately picked as the optimal rotation angle  $\theta^*$ . The global voting algorithm for 1-DOF rotation estimation is outlined in Algorithm 5.

Overall, the 4-DOF correspondence-based point cloud registration problem can be easily solved by running the 3-DOF translation search in Algorithm 4 and the 1-DOF rotation estimation in Algorithm 5 sequentially. Due to the decoupling of the joint pose, the proposed method is theoretically more efficient than the existing method [Cai+19]. It should be noted that the separately optimal solutions of these two sub-problems do not necessarily result in the globally optimal solution for the raw 4-DOF problem. Since the objective of the first sub-problem is formulated as a problem only about translation, the constraint about rotation is dropped. It is like dropping rank-one constraint in semidefinite programming (SDP) relaxation [YSC20]. Therefore, the decomposition is, to some extent, a relaxation, implying that the two sub-problems are not exactly the same as the raw 4-DOF problem. Nonetheless, the proposed two-stage method is deterministic, robust, and efficient, verified by extensive experiments in Section 5.4.1.

### Simultaneous Pose and Correspondence Registration

This section extends the proposed method to solve the challenging registration case, where the putative correspondences are unknown. However, all possible correspondences between

---

**Algorithm 5:** Global voting for 1-DOF rotation estimation (solution to problem (5.27))

---

**Input:** Globally optimal solution  $\mathbf{t}^*$ , Set of candidate correspondences

$$\mathcal{K} = \{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^N, \text{ and Gravity directions } \mathbf{v}_p, \mathbf{v}_q.$$

**Output:** Globally optimal solution  $\theta^*$ .

- 1 Calculate  $\mathbf{R}_{\mathbf{v}_p}^{\mathbf{v}_q}$  according to Eq. (5.30);
  - 2 Calculate angle  $\theta_i$ ,  $i = 1, \dots, N$  for each candidate correspondence in  $\mathcal{K}$ ;
  - 3 Operate histogram voting in the interval  $[-\pi, \pi]$  for angle  $\theta_i$ ,  $i = 1, \dots, N$ ;
- 

the source point cloud  $\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^N$  and target point cloud  $\mathcal{Q} = \{\mathbf{q}_j\}_{j=1}^M$  can be generated according to the hypothesis of all-to-all correspondences. Consequently, the objective function (5.9) can be modified to:

$$S(\mathbf{t}|\mathcal{K}, \zeta) = \sum_{i=1}^N \sum_{j=1}^M \mathbb{I}(|\angle(\mathbf{p}_i + \mathbf{t}, \mathbf{v}_p) - \angle(\mathbf{q}_j, \mathbf{v}_q)| \leq \zeta). \quad (5.31)$$

The modified upper and lower bounds for any translation sub-branch  $\mathbb{B}$  are:

$$\bar{S}(\mathbb{B}) = \sum_{i=1}^N \sum_{j=1}^M \mathbb{I}(|\alpha_i^0 - \angle(\mathbf{q}_j, \mathbf{v}_q)| - r_i \leq \zeta), \quad (5.32)$$

$$\underline{S}(\mathbb{B}) = \sum_{i=1}^N \sum_{j=1}^M \mathbb{I}(|\alpha_i^0 - \angle(\mathbf{q}_j, \mathbf{v}_q)| \leq \zeta). \quad (5.33)$$

Thus the globally optimal solution  $\mathbf{t}^*$  can also be obtained by Algorithm 4, i.e.,

$$\mathbf{t}^* = \arg \max_{\mathbf{t} \in \mathbb{R}^3} S(\mathbf{t}|\mathcal{K}, \zeta). \quad (5.34)$$

The next sub-problem of estimating the rotation  $\theta^*$  can also be solved by Algorithm 5, since the first step implicitly estimates the correspondences when estimating the optimal  $\mathbf{t}^*$ . Under the modified formulation, the proposed 4-DOF correspondence-based method could be easily extended to solve the challenging SPCR problem, including the correspondence-free translation search and correspondence-based rotation estimation. The performance of the extended method is presented in Section 5.4.1.

### 5.3 Transformation Decoupling by Screw Theory

In this section, we reformulate the registration problem by *screw theory* [Bal76] and propose a novel transformation decoupling strategy accordingly. This strategy decouples the original 4-DOF registration problem into three sub-problems with 1-DOF, 2-DOF, and 1-DOF. Concretely, the first 1-DOF refers to the translation along the rotation axis, and we propose a polynomial-time method to solve it based on interval stabbing. The second 2-DOF represents the pole which is an auxiliary variable in screw theory. We reformulate it as a linear model fit-

ting problem by exploring geometric constraints, and utilize a BnB-based algorithm to search for the globally optimal solution. The last 1-DOF refers to the rotation angle, and we propose an efficient global voting method for its estimation. The proposed three-stage search method sequentially solves three sub-problems in low-dimensional spaces to achieve efficient and deterministic registration. It fills the gap between efficient but non-deterministic heuristics (e.g., RANSAC) and deterministic but exhaustive global algorithms (e.g., BnB). We explore the scalability of the proposed method through extreme synthetic experiments. For instance, the proposed method accurately solves the registration problem with 95% outlier rate and  $10^5$  correspondences in 0.1 seconds (see Table 5.3), while one of the state-of-the-art (SOTA) deterministic global methods, FMP+BnB [Cai+19], takes about 648 seconds. In another more extreme case, the proposed method solves the registration problem with 95% outlier rate and  $10^6$  correspondences in 2 seconds, while another SOTA method GROR [Yan+22a] takes at least 1800 seconds.

In conclusion, the main contributions are as follows:

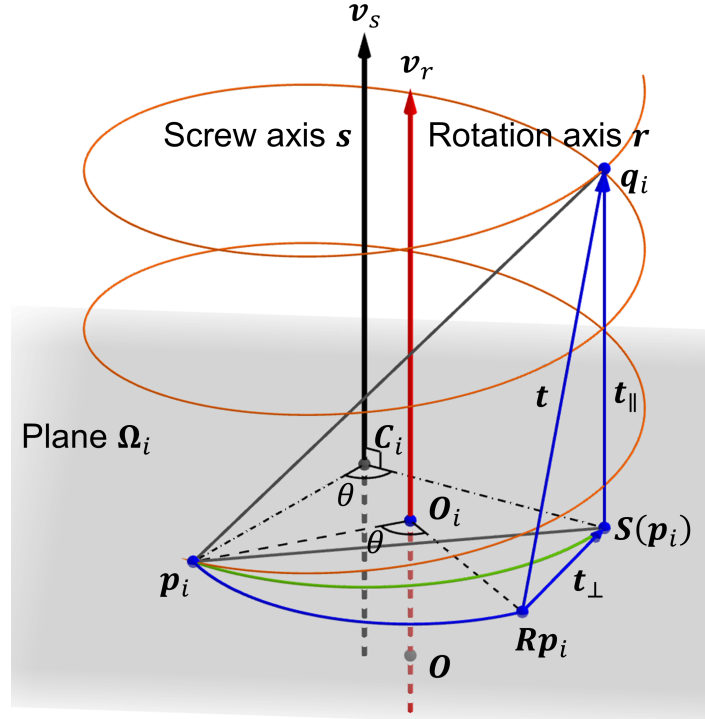
- By reformulating the point cloud registration problem from the perspective of screw theory, this section accordingly proposes a novel transformation decoupling strategy with the aid of known gravity directions. It decouples the 4-DOF registration problem into three sub-problems with 1-DOF, 2-DOF, and 1-DOF, respectively, significantly improving the algorithm efficiency.
- To achieve highly robust registration, this section proposes an efficient and deterministic three-stage search strategy for the decoupled sub-problems, which contains interval stabbing, BnB, and global voting techniques. The proposed method has a good robustness-efficiency trade-off.
- The proposed method is extended to solve the challenging SPCR problem without the all-to-all correspondence assumption. It avoids the potential hard combinatorial problem when input point clouds are of large size and thus is highly efficient.

The rest of this section is organized as follows: Section 5.3.1 illustrates the problem formulation of the 4-DOF point cloud registration. Section 5.3.2 demonstrates the principle and details the proposed registration method. Section 5.3.3 introduces the extended method for the correspondence-free registration problem.

### 5.3.1 Problem Formulation

#### Inlier Set Maximization with Screw Theory

Typically, outlier-robust rigid point cloud registration can be formulated as a consensus maximization (a.k.a. inlier set maximization) problem, where the cardinality of the inlier set is to be maximized. We assume that the set of putative correspondences  $\mathcal{C} = \{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^N$  is extracted between the source point cloud  $\mathcal{P}$  and the target point cloud  $\mathcal{Q}$ , where  $\mathbf{p}_i, \mathbf{q}_i \in \mathbb{R}^3$ . Considering the existence of noise, we introduce a threshold  $\epsilon$  for the identification of inliers.



**Figure 5.6:** Schematic of Chasles' Theorem and screw transformation. Specifically, the Euclidean transformation of point  $p_i$ , i.e.,  $Rp_i + t$  can be represented by a screw rotation  $S(p_i)$  combined with a screw translation  $t_{\parallel}$ . The screw axis is defined by a unique point  $C_i$  in the plane  $\Omega_i$  and the rotation axis  $r$ .

Thus, the inlier set maximization problem for rigid registration is commonly organized as below,

$$\begin{aligned} \max_{R, t, \mathcal{I} \subseteq \mathcal{H}} |\mathcal{I}| \\ \text{s.t. } \|Rp_i + t - q_i\| \leq \epsilon, \forall i \in \mathcal{I}, \end{aligned} \quad (5.35)$$

where  $R \in \mathbb{SO}(3)$  is the rotation matrix,  $t \in \mathbb{R}^3$  is the translation vector,  $\mathcal{H} = \{1, \dots, N\}$  is the set of indices for  $\mathcal{C}$ ,  $\mathcal{I}$  represents the inlier set,  $|\cdot|$  is the cardinality of a set, and  $\|\cdot\|$  is the  $L_2$ -norm. This optimization problem aims to find the optimal  $R^*$  and  $t^*$  to maximize the cardinality of the inlier set. This problem is inherently a 6-DOF optimization problem since both  $R$  and  $t$  require three parameters to be defined.

*Screw theory* [Bal76] is a widely used important tool in robot mechanics [LP17] and computational geometry [De +97]. It includes a fundamental theorem, which is known as *Chasles' Theorem* [HH78; BR90], as shown in Theorem 5.1. In this section, we will reformulate the rigid point cloud registration problem from the perspective of screw theory.

**Theorem 5.1** (Chasles' Theorem [HH78; BR90]). *Each Euclidean transformation in three-dimensional space has a screw axis, and the transformation can be decomposed into a rotation about and a translation along this screw axis.*

Chasles' Theorem indicates that the six parameters of a Euclidean transformation contain the four independent components that define the *screw axis*, together with the rotation angle about and translation along this screw axis [Bal76]. Typically, the six parameters defining the Euclidean transformation can also be given by three Euler angles of rotation and three

translation components. In screw theory, the Euclidean transformation is also known as the *screw transformation*, the rotation about the screw axis is known as the *screw rotation*, and the translation along the screw axis is known as the *screw translation*. Mathematically, according to Chasles' Theorem, the screw transformation of  $\mathbf{p}_i$  can be represented by

$$\mathbf{T}(\mathbf{p}_i) = \mathbf{R}\mathbf{p}_i + \mathbf{t} \quad (5.36a)$$

$$= (\mathbf{R}\mathbf{p}_i + \mathbf{t}_\perp) + \mathbf{t}_\parallel \quad (5.36b)$$

$$= \mathbf{S}(\mathbf{p}_i) + \mathbf{t}_\parallel, \quad (5.36c)$$

where  $\mathbf{T} \in \mathbb{SE}(3)$  denotes the Euclidean transformation,  $\mathbf{S}(\cdot)$  denotes the screw rotation, and  $\mathbf{t}_\perp$  and  $\mathbf{t}_\parallel$  are the translation components perpendicular and parallel to the screw axis, respectively. Therefore, the screw rotation  $\mathbf{S}(\cdot)$  contains the rotation  $\mathbf{R}$  and the translation  $\mathbf{t}_\perp$ . The screw translation is the translation component  $\mathbf{t}_\parallel$ . The related geometrical interpretation of Chasles' Theorem and screw transformation is shown in Fig. 5.6. We define the unit-norm constrained orientation vectors of the screw and rotation axis as  $\mathbf{v}_s$  and  $\mathbf{v}_r$ . Specifically, the screw axis  $\mathbf{s}$  can be defined by the orientation vector  $\mathbf{v}_r$  and a unique point  $\mathbf{C}_i \in \mathbb{R}^3$  in the *rotation plane*  $\Omega_i$  that is perpendicular to  $\mathbf{v}_r$  and through the point  $\mathbf{p}_i$ . Thus the screw axis is parallel to the rotation axis, i.e.,  $\mathbf{v}_s$  is parallel to  $\mathbf{v}_r$ . In three-dimensional space, the screw rotation  $\mathbf{S}(\cdot)$  is then defined as

$$\mathbf{S}(\mathbf{p}_i) = \mathbf{R}(\mathbf{p}_i - \mathbf{C}_i) + \mathbf{C}_i. \quad (5.37)$$

Notably, after screw rotation, the transformed point  $\mathbf{S}(\mathbf{p}_i)$  remains in the rotation plane that is perpendicular to  $\mathbf{v}_s$  and  $\mathbf{v}_r$ . For this unique point  $\mathbf{C}_i$ , we can obtain

$$\mathbf{S}(\mathbf{C}_i) = \mathbf{C}_i. \quad (5.38)$$

Therefore, according to the definition of rotation, the rigid motion  $\mathbf{S}(\cdot)$ , which can keep a point fixed (i.e.,  $\mathbf{C}_i$ ), is a rotational motion called *screw rotation*. Both the orientation of the screw axis and the rotation angle of the screw rotation are identical to that of the original rotation  $\mathbf{R}$ . The only difference is that the rotation axis  $\mathbf{r}$  is through the origin  $\mathbf{O}$  by default, while the screw axis  $\mathbf{s}$  is through the point  $\mathbf{C}_i$ . In general, the 6-DOF rigid registration problem (5.35) under screw theory can be rewritten as

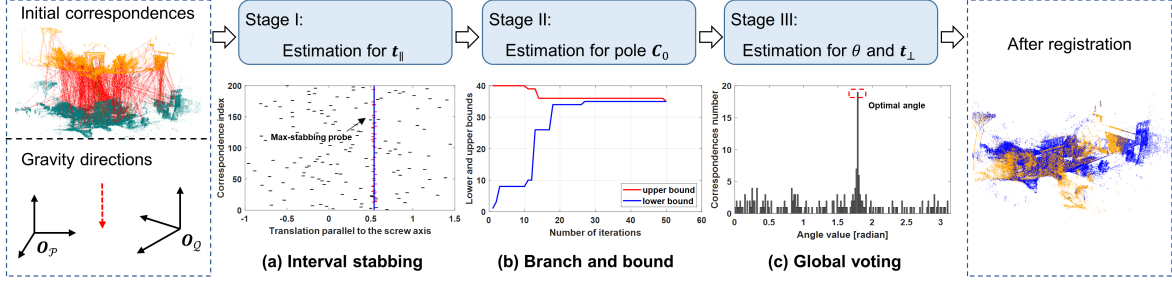
$$\begin{aligned} & \max_{\mathbf{R}, \mathbf{C}_i, \mathbf{t}_\parallel, \mathcal{I} \subseteq \mathcal{H}} |\mathcal{I}| \\ & \text{s.t.} \quad \|\mathbf{S}(\mathbf{p}_i) + \mathbf{t}_\parallel - \mathbf{q}_i\| \leq \epsilon, \quad \forall i \in \mathcal{I}. \end{aligned} \quad (5.39)$$

Solving this optimization problem thus amounts to searching for  $\mathbf{R}$  (a 3-DOF problem),  $\mathbf{C}_i$  (a 2-DOF problem), and  $\mathbf{t}_\parallel$  (a 1-DOF problem).

On the other hand, we can also obtain the following theorem about the planar specialization of Chasles' Theorem.

**Theorem 5.2** (Planar Chasles' Theorem [HH78; BR90]). *When a Euclidean transformation specializes to a planar transformation in two-dimensional space, the screw axis becomes a pole,*





**Figure 5.7:** Calculation flow of the proposed three-stage method for correspondence-based registration. The source, target, and aligned point clouds are green, orange, and blue, respectively. In the initial correspondences set, green and red line segments represent inliers and outliers, respectively (only a subset of correspondences is shown for visual clarity). For the interval stabbing part, the candidate intervals are represented by the black line segments, while the intervals crossed by the max-stabbing probe are depicted as the red line segments.

and the planar transformation can be represented by a rotation about this pole.

Theorem 5.2 indicates that a 2D rigid transformation (e.g., transform point  $p$  to  $q$ ) in the plane can be converted to a pure rotational motion around a unique pole  $C_0$ . This pure rotational motion is a 2D screw rotation, denoted by  $S_2(\cdot)$ , and we can obtain

$$S_2(p) = R_2(p - C_0) + C_0 = q, \quad (5.40)$$

where  $p, q, C_0 \in \mathbb{R}^2$ , and  $R_2$  is a 2D rotation matrix. In this study, Theorem 5.2 is also one of the fundamental parts of the proposed transformation decoupling strategy.

### Rigid Registration with Gravity Prior

In the case that the gravity directions are given, we assume they are denoted by unit vectors  $v_p, v_q$  for the source and target point clouds  $\mathcal{P}, \mathcal{Q}$ , respectively. The constraint of gravity directions is given by

$$v_q = R v_p. \quad (5.41)$$

The solution of this equation is given by [BC17; LCK23]

$$R = R(\theta, v_q) R_{v_p}^{v_q}, \quad (5.42)$$

where  $R(\theta, v_q)$  is the rotation that rotates  $\theta$  about axis  $v_q$ , and  $R_{v_p}^{v_q}$  is the rotation that rotates  $v_p$  to  $v_q$  with the minimum geodesic motion. Eq. (5.42) indicates that the rotation  $R$  is only dependent on the rotation angle  $\theta \in [-\pi, \pi]$  when the gravity directions are prior known. Without loss of generality, we can align the Z-axis of the source point cloud  $\mathcal{P}$  to the Z-axis of the target point cloud  $\mathcal{Q}$  by the following rotation,

$$p'_i = R_{v_p}^{v_q} p_i. \quad (5.43)$$

After this rotation, the 6-DOF rigid registration problem (5.39) can be reduced to a 4-DOF problem, i.e.,

$$\begin{aligned} & \max_{\theta, \mathbf{C}, \mathbf{t}_{\parallel}, \mathcal{I} \subseteq \mathcal{H}} |\mathcal{I}| \\ & \text{s.t.} \quad \|\mathbf{S}'(\mathbf{p}'_i) + \mathbf{t}_{\parallel} - \mathbf{q}_i\| \leq \epsilon, \quad \forall i \in \mathcal{I}, \end{aligned} \quad (5.44)$$

where  $\mathbf{S}'(\mathbf{p}'_i) = \mathbf{R}'(\mathbf{p}'_i - \mathbf{C}) + \mathbf{C}$ , and we set  $\mathbf{R}' \triangleq \mathbf{R}(\theta, \mathbf{v}_q)$  for convenience. Please note that in problem (5.44), the orientation of the screw axis is parallel to  $\mathbf{v}_q$  rather than the original rotation axis  $\mathbf{v}_r$ , i.e.,  $\mathbf{v}_q$  is the current rotation axis. Accordingly,  $\mathbf{t}_{\parallel}$  in problem (5.44) is the translation component parallel to  $\mathbf{v}_q$ , and  $\mathbf{t}_{\perp}$  is the component perpendicular to  $\mathbf{v}_q$ .

### 5.3.2 Three-Stage Consensus Maximization Registration

Due to the high-dimensional parameter space, solving the 4-DOF problem (5.44) jointly is relatively time-consuming [Cai+19]. To accelerate the registration, we first reduce the 4-DOF original problem to a 1-DOF sub-problem with the aid of known gravity directions in Stage I. The 1-DOF sub-problem is estimating the translation parallel to the screw axis. Then we decouple the remaining 3-DOF sub-problem into a 2-DOF and a 1-DOF sub-problem by screw theory, which is addressed in Stage II and Stage III, respectively. The 2-DOF sub-problem is searching for the *pole*, an auxiliary variable in screw theory. The last 1-DOF sub-problem is estimating the rotation angle. After acquiring the rotation angle, we can readily calculate the translation orthogonal to the rotation axis and thereby obtain the final optimal solution. Similar to the original problem, we formulate all three sub-problems as consensus maximization problems. The pipeline of the proposed method is given in Fig. 5.7.

#### Stage I: Estimation for the Translation Parallel to the Screw Axis

Using the known gravity directions, we can reduce the original problem into a 1-DOF sub-problem that solely involves the translation parallel to the screw axis. Specifically, for an ideal inlier correspondence  $(\mathbf{p}_i, \mathbf{q}_i)$ , we have the following derivation about the original constraint in Eq. (5.44),

$$\mathbf{q}_i = \mathbf{S}'(\mathbf{p}'_i) + \mathbf{t}_{\parallel} \quad (5.45a)$$

$$\Leftrightarrow \mathbf{q}_i - \mathbf{p}'_i = \mathbf{S}'(\mathbf{p}'_i) - \mathbf{p}'_i + \mathbf{t}_{\parallel} \quad (5.45b)$$

$$\Rightarrow \mathbf{v}_q^T (\mathbf{q}_i - \mathbf{p}'_i) = \mathbf{v}_q^T [\mathbf{S}'(\mathbf{p}'_i) - \mathbf{p}'_i + \mathbf{t}_{\parallel}] \quad (5.45c)$$

$$\Leftrightarrow \mathbf{v}_q^T (\mathbf{q}_i - \mathbf{p}'_i) = \|\mathbf{t}_{\parallel}\| \quad (5.45d)$$

where Eq. (5.45d) is from the fact that vector  $\mathbf{S}'(\mathbf{p}'_i) - \mathbf{p}'_i$  is perpendicular to  $\mathbf{v}_q$  (see Fig. 5.6), and  $\mathbf{t}_{\parallel}$  is parallel to  $\mathbf{v}_q$ . Considering the noise, we have the following new inlier constraint,

$$\left| \|\mathbf{t}_{\parallel}\| + \mathbf{v}_q^T (\mathbf{p}'_i - \mathbf{q}_i) \right| \leq \delta, \quad (5.46)$$

where  $\delta$  is the inlier threshold. Since the gravity direction  $\mathbf{v}_q$  is given, the constraint in Eq. (5.46) is only dependent on  $\|\mathbf{t}_{\parallel}\|$ . Accordingly, we can rewrite Eq. (5.46) in the form of

an interval for  $l \triangleq \|\mathbf{t}_{\parallel}\|$ , i.e.,

$$-\delta - \mathbf{v}_q^T(\mathbf{p}'_i - \mathbf{q}_i) \leq l \leq \delta - \mathbf{v}_q^T(\mathbf{p}'_i - \mathbf{q}_i). \quad (5.47)$$

Given the set of initial correspondences and gravity directions, the first sub-problem aims at estimating the optimal  $l^*$  to maximize the cardinality of the inlier set, which is defined by

$$\begin{aligned} & \max_{l, \mathcal{I}_1 \subseteq \mathcal{H}} |\mathcal{I}_1| \\ \text{s.t. } & l \in [l_i^-, l_i^+], \forall i \in \mathcal{I}_1, \\ & l_i^- = -\delta - \mathbf{v}_q^T(\mathbf{p}'_i - \mathbf{q}_i), \\ & l_i^+ = \delta - \mathbf{v}_q^T(\mathbf{p}'_i - \mathbf{q}_i), \end{aligned} \quad (5.48)$$

where  $\mathcal{I}_1$  denotes the inlier set that only satisfies the constraint in Eq. (5.48). From the perspective of computational geometry, solving the consensus maximization problem (5.48) is a typical *interval stabbing* problem [De +97]. Recently, the interval stabbing technology has been widely used to solve different geometric optimization problems [BC17; Cai+19; PTV22; Yan+22a; Zha+23a].

As depicted in Fig. 5.7, the interval stabbing problem is concerned with finding a probe (represented by the blue line segment) that stabs the maximum number of intervals. The interval stabbing problem can be efficiently solved with a time complexity of  $\mathcal{O}(N \log N)$ . In contrast to existing algorithms, the proposed interval stabbing algorithm (Algorithm 6) enhances the precision by returning the midpoint of the maximum overlapping interval instead of the common left endpoint. The *max-stabbing number* returned from Algorithm 6 is the maximized cardinality of the inlier set. The returned *max-stabbing position* is the value of optimal  $l^*$ . Besides, since  $l^*$  is a scalar, we can obtain the optimal translation parallel to  $\mathbf{v}_q$  by

$$\mathbf{t}_{\parallel}^* = l^* \cdot \mathbf{v}_q \quad (5.49)$$

### Stage II: Searching for the Pole

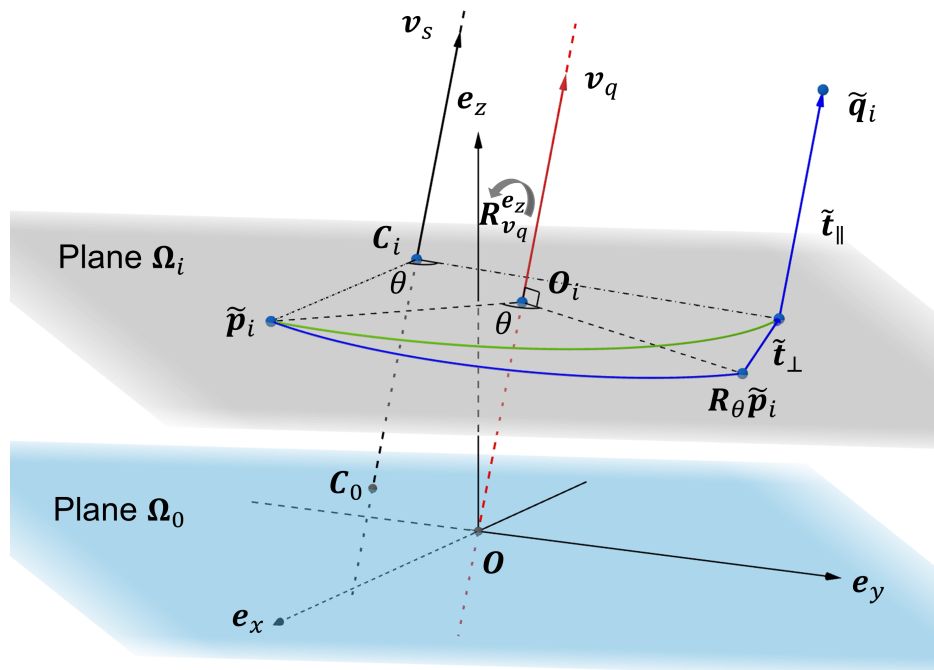
In this section, we first reduce the original 4-DOF problem (Eq. (5.44)) into a 2D-2D rigid registration problem with 3-DOF using *plane projection*. Subsequently, we decouple this 3-DOF registration problem into a 2-DOF sub-problem and a 1-DOF sub-problem by screw theory. This section focuses on solving the 2-DOF sub-problem, concretely searching for the pole.

**Plane Projection and Pole** For an ideal inlier correspondence  $(\mathbf{p}_i, \mathbf{q}_i)$ , the constraint in Eq. (5.44) can be rewritten as,

$$\mathbf{q}_i = \mathbf{R}' \mathbf{p}'_i + \mathbf{t}_{\perp} + \mathbf{t}_{\parallel}. \quad (5.50)$$

**Algorithm 6:** Interval stabbing algorithm (solution to problem (5.48))**Input:** Intervals  $\{[l_i^-, l_i^+]\}_{i=1}^N$ **Output:** Optimal  $l^*$ 

- 1 Sort the left sides  $\{l_i^-\}_{i=1}^N$  in ascending order as  $\{l_j^-\}_{j=1}^N$ ;
- 2 Sort the right sides  $\{l_i^+\}_{i=1}^N$  in ascending order as  $\{l_k^+\}_{k=1}^N$ ;
- 3 Initialize the number of stabbed intervals  $n = 0$ , and the max-stabbing number  $max = 0$ ;
- 4 Initialize the index  $j = 1, k = 1$ ;
- 5 **while**  $j \leq N$  **AND**  $k \leq N$  **do**
- 6     **if**  $l_j^- \leq l_k^+$  **then**
- 7          $n = n + 1$ ;
- 8         **if**  $n > max$  **then**
- 9              $max = n, \alpha = j, \beta = k$ ;
- 10         **end**
- 11          $j = j + 1$ ;
- 12     **else**
- 13          $n = n - 1, k = k + 1$ ;
- 14     **end**
- 15 **end**
- 16  $l^* = (l_\alpha^- + l_\beta^+)/2$

**Figure 5.8:** The geometrical interpretation of the coordinate conversion by  $R_{v_q}^{e_z}$  and the plane projection to  $\Omega_0$ .

It can be observed that only the translation  $\mathbf{t}_{\parallel}$  is along the screw axis, while the rotation  $\mathbf{R}'$  and the translation  $\mathbf{t}_{\perp}$  maintain the point  $\mathbf{p}'_i$  within the rotation plane  $\Omega_i$ . This plane is perpendicular to  $\mathbf{v}_q$  and passes through  $\mathbf{p}'_i$ . Therefore, given the gravity direction  $\mathbf{v}_q$ , we can project the set of correspondence  $\mathcal{C}$  from the three-dimensional space onto a two-dimensional plane  $\Omega_0$  that is perpendicular to  $\mathbf{v}_q$  and through the origin  $\mathbf{O}$  (i.e.,  $\Omega_0$  is parallel to each  $\Omega_i$ ). After plane projection, the original 3D rigid transformation can be reduced to a 2D transformation with 3-DOF, which solely contains the 1-DOF rotation angle  $\theta$  and the 2-DOF translation  $\mathbf{t}_{\perp}$ . Interestingly, the 2D rigid transformation can be represented by a 2D screw rotation (Theorem 5.2). In the next step, we first conduct the following coordinate conversion operation to achieve plane projection, and the geometrical illustration is given in Fig. 5.8.

Firstly, we introduce a new coordinate system, in which the rotation axis  $\mathbf{v}_q$  corresponds to the Z-axis (denoted by  $\mathbf{e}_z = [0, 0, 1]^T$ ). The transformation from the original coordinate system to the new one is given by

$$\mathbf{e}_z = \mathbf{R}_{\mathbf{v}_q}^{e_z} \mathbf{v}_q. \quad (5.51)$$

We can transform each vector in Eq. (5.50) to the new coordinate system by

$$\tilde{\mathbf{p}}_i = \mathbf{R}_{\mathbf{v}_q}^{e_z} \mathbf{p}'_i, \quad \tilde{\mathbf{q}}_i = \mathbf{R}_{\mathbf{v}_q}^{e_z} \mathbf{q}_i, \quad \tilde{\mathbf{t}}_{\perp} = \mathbf{R}_{\mathbf{v}_q}^{e_z} \mathbf{t}_{\perp}, \quad \tilde{\mathbf{t}}_{\parallel} = \mathbf{R}_{\mathbf{v}_q}^{e_z} \mathbf{t}_{\parallel}. \quad (5.52)$$

Then we have the following coordinate conversion,

$$\mathbf{q}_i = \mathbf{R}' \mathbf{p}'_i + \mathbf{t}_{\perp} + \mathbf{t}_{\parallel} \quad (5.53a)$$

$$\Leftrightarrow \mathbf{R}_{\mathbf{v}_q}^{e_z} \mathbf{q}_i = \mathbf{R}_{\mathbf{v}_q}^{e_z} \mathbf{R}' (\mathbf{R}_{\mathbf{v}_q}^{e_z})^T \mathbf{R}_{\mathbf{v}_q}^{e_z} \mathbf{p}'_i + \mathbf{R}_{\mathbf{v}_q}^{e_z} \mathbf{t}_{\perp} + \mathbf{R}_{\mathbf{v}_q}^{e_z} \mathbf{t}_{\parallel} \quad (5.53b)$$

$$\Leftrightarrow \tilde{\mathbf{q}}_i = \mathbf{R}_{\theta} \tilde{\mathbf{p}}_i + \tilde{\mathbf{t}}_{\perp} + \tilde{\mathbf{t}}_{\parallel}, \quad (5.53c)$$

where  $\mathbf{R}_{\theta} \triangleq \mathbf{R}_{\mathbf{v}_q}^{e_z} \mathbf{R}' (\mathbf{R}_{\mathbf{v}_q}^{e_z})^T$  is the rotation matrix in the new coordinate system, as shown in Fig. 5.8. The rotation axis in this coordinate system is now the Z-axis, thus  $\mathbf{R}_{\theta}$  can be denoted as

$$\mathbf{R}_{\theta} \triangleq \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (5.54)$$

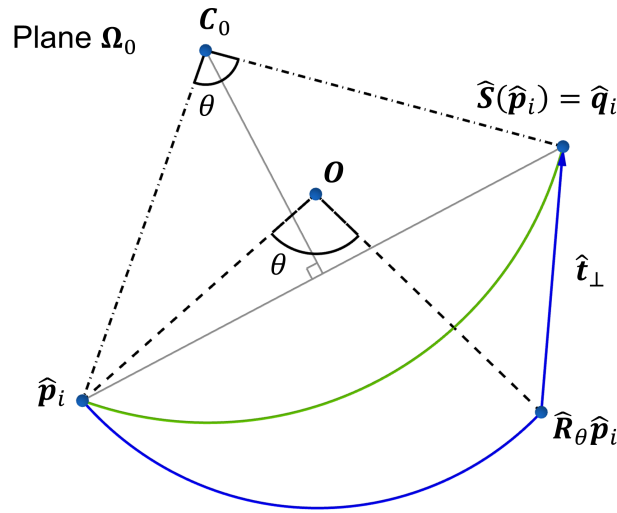
Furthermore,  $\tilde{\mathbf{t}}_{\perp}$  and  $\tilde{\mathbf{t}}_{\parallel}$  have the following form

$$\tilde{\mathbf{t}}_{\perp} \triangleq [t_a, t_b, 0]^T, \quad \tilde{\mathbf{t}}_{\parallel} \triangleq [0, 0, l]^T. \quad (5.55)$$

After projecting the 3D rigid transformation in Eq. (5.53c) onto the 2D plane  $\Omega_0$ , we can obtain the 2D rigid transformation as shown below

$$\hat{\mathbf{q}}_i = \hat{\mathbf{R}}_{\theta} \hat{\mathbf{p}}_i + \hat{\mathbf{t}}_{\perp}, \quad \hat{\mathbf{R}}_{\theta} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \quad (5.56)$$

where  $\hat{\mathbf{t}}_{\perp} = [t_a, t_b]^T$ ,  $\hat{\mathbf{p}}_i \triangleq [x_p, y_p]^T$ , and  $\hat{\mathbf{q}}_i \triangleq [x_q, y_q]^T$ . The geometrical illustration for plane projection is given in Fig. 5.9. Estimating the optimal  $\hat{\mathbf{R}}_{\theta}$  and  $\hat{\mathbf{t}}_{\perp}$  is a typical 2D-2D rigid



**Figure 5.9:** The geometrical interpretation of the 2D rigid transformation and the 2D screw rotation after plane projection.

registration problem.

In the formulation based on screw theory, the 2D rigid transformation can be reformulated by a 2D screw rotation, as presented in Theorem 5.2. The 2D screw rotation is a pure rotational motion around *pole*  $C_0 \in \mathbb{R}^2$ , which is the projection of  $C_i \in \mathbb{R}^3$ . We can rewrite Eq. (5.56) as,

$$\hat{q}_i = \hat{R}_\theta(\hat{p}_i - C_0) + C_0 = \hat{S}(\hat{p}_i). \quad (5.57)$$

Accordingly, under screw theory, the 2D-2D rigid registration problem consists of estimating the 2-DOF rotation center  $C_0$  and the 1-DOF rotation angle  $\theta$ , as shown in Fig. 5.9. An essential property of this rotation center (pole) is that it must stand on the vertical bisector of each line segment  $\hat{p}_i \hat{q}_i$ , as presented in Proposition 5.2.

**Proposition 5.2.** *When a point rotates around a rotation center on a plane, the rotation center must fall on the perpendicular bisector of the line segment connecting the two corresponding points.*

We first define the perpendicular bisector as a linear function of the form  $a_i x + b_i y + c_i = 0$ . For an ideal inlier correspondence  $(\hat{p}_i, \hat{q}_i)$ , we can obtain

$$\begin{aligned} a_i &= x_q - x_p, \\ b_i &= y_q - y_p, \\ c_i &= -\left(\frac{x_p + x_q}{2} a_i + \frac{y_p + y_q}{2} b_i\right). \end{aligned} \quad (5.58)$$

Pole  $C_0 \triangleq [C_x, C_y]^T$  should stand on this line. However, when the rotation angle tends to zero,  $C_0$  stands in an infinite far position, which can not be expressed in a normal way. To avoid this limitation, we use the homogeneous coordinate  $\hat{C} \triangleq [c_x, c_y, w]^T$  to represent pole  $C_0$ , and the constraint of such coordinate is  $\|\hat{C}\| = 1$ . We also use the vector expression for

each perpendicular bisector, i.e.,  $\mathbf{n}_i \triangleq [a_i, b_i, c_i]^T$ . Therefore, the constraint of the pole can be expressed as

$$a_i c_x + b_i c_y + c_i w = \mathbf{n}_i^T \hat{\mathbf{C}} = 0. \quad (5.59)$$

Considering the noise, estimating pole  $\mathbf{C}_0$  can be formulated as a *linear model fitting problem* [Wan+21b; Liu+22]. We can solve this problem by addressing the following consensus maximization problem, in which we seek an optimal unit-norm constrained  $\hat{\mathbf{C}}$  with the largest number of candidate inliers.

$$\begin{aligned} \max_{\hat{\mathbf{C}}, \mathcal{I}_2 \subseteq \mathcal{I}_1^*} & |\mathcal{I}_2| \\ \text{s.t.} & |\mathbf{n}_i^T \hat{\mathbf{C}}| \leq \tau, \quad \forall i \in \mathcal{I}_2, \\ & \|\hat{\mathbf{C}}\| = 1, \end{aligned} \quad (5.60)$$

where  $\mathcal{I}_2$  is the inlier set extracted from the cardinality-maximized inlier set  $\mathcal{I}_1^*$ , and  $\tau$  is the inlier threshold.

**Branch and Bound** Since searching for the exact solution of the consensus maximization problem (5.60) is NP-hard [TZN20], we utilize the globally optimal and deterministic BnB algorithm [Mor+16; Sch11] to solve this optimization problem. Specifically, the BnB algorithm systematically explores the entire parameter space (a.k.a. solution domain) by iteratively dividing it into smaller sub-branches and calculating the upper and lower bound for each sub-branch. It discards those sub-branches where larger objective function values than the current optimal value are impossible. As the solution domain progressively narrows down, the gap between the upper and lower bounds gradually diminishes until zero, and then the BnB algorithm achieves the optimal solution. An example of the convergence curve of the proposed BnB algorithm is given in Fig. 5.7.

The first step in constructing the BnB algorithm is the parameterization of the solution domain. Geometrically, the unit-norm constrained vector  $\hat{\mathbf{C}}$  lies on the surface of a *unit sphere*. We denote the unit sphere as  $\mathbb{S}^2$ . Since  $\hat{\mathbf{C}}$  and  $-\hat{\mathbf{C}}$  have the same inlier set, we can set the solution domain of  $\hat{\mathbf{C}}$  as a hemisphere denoted by  $\mathbb{S}^{2+} = \{\mathbf{h} = [h_1, h_2, h_3]^T \mid \|\mathbf{h}\| = 1, h_3 \geq 0\}$ . We then use a compact representation method, *exponential mapping* [HK09; LCK20b], to map the 3D unit hemisphere to a 2D disk. Concretely, the vector  $\mathbf{h} \in \mathbb{S}^{2+}$  corresponds to a unique point  $\varphi \in \mathbb{R}^2$  in the disk,

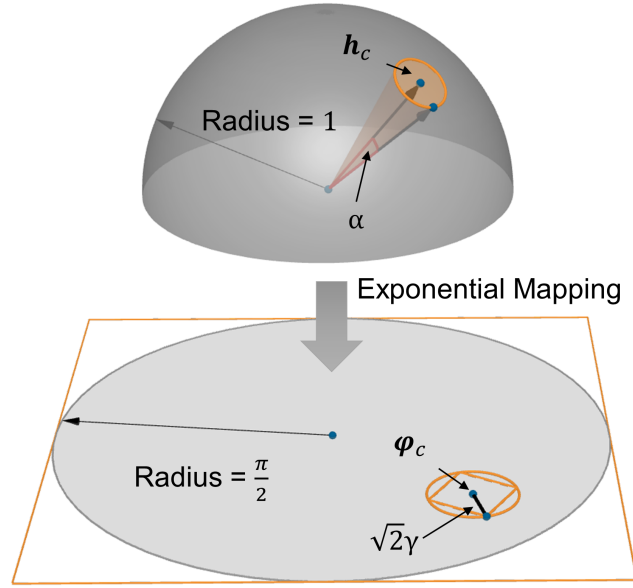
$$\mathbf{h} = [\sin \omega \cdot \hat{\varphi}^T, \cos \omega]^T \quad (5.61a)$$

$$\varphi = \hat{\varphi} \cdot \omega \quad (5.61b)$$

where  $\omega \in [0, \pi/2]$ , and  $\hat{\varphi}$  is a unit vector in  $\mathbb{R}^2$ . Therefore, to facilitate manipulation, the solution domain of  $\hat{\mathbf{C}}$  is defined as a circumscribed square (with a radius of  $\pi/2$ ) of the disk domain, as shown in Fig. 5.10.

The next step is estimating the upper and lower bound for the sub-branch. We first introduce the following lemma.

**Lemma 5.1.** *Given a square-shaped sub-branch  $\mathbb{B}$  in the exponential mapping plane, its center*



**Figure 5.10:** Exponential mapping of the solution domain and the geometrical interpretation of Lemma 5.1.

is  $\varphi_c \in \mathbb{R}^2$  and half-side length is  $\gamma$ . For  $\forall \varphi \in \mathbb{B}$ , we have

$$\angle(\mathbf{h}, \mathbf{h}_c) \leq \|\varphi - \varphi_c\| \leq \sqrt{2}\gamma, \quad (5.62)$$

where  $\varphi$  and  $\varphi_c$  correspond to  $\mathbf{h} \in \mathbb{S}^{2+}$  and  $\mathbf{h}_c \in \mathbb{S}^{2+}$ , respectively.

*Proof.* The completed proof is given in [LCK20b].  $\square$

Based on Lemma 5.1, the upper and lower bound of the proposed BnB algorithm for problem (5.60) can be set as

**Proposition 5.3.** Given a square-shaped sub-branch  $\mathbb{B}$ , whose center is  $\varphi_c \in \mathbb{R}^2$  (corresponds to  $\mathbf{h}_c \in \mathbb{S}^{2+}$  by exponential mapping) and half-side length is  $\gamma$ , the upper bound  $U(\mathbb{B})$  and lower bound  $L(\mathbb{B})$  can be set as

$$U(\mathbb{B}) = \sum_{i=1}^{N'} \mathbb{I}(|\mathbf{n}_i^T \mathbf{h}_c| \leq \Psi_i), \quad (5.63a)$$

$$L(\mathbb{B}) = \sum_{i=1}^{N'} \mathbb{I}(|\mathbf{n}_i^T \mathbf{h}_c| \leq \tau), \quad (5.63b)$$

$$\Psi_i = \begin{cases} \|\mathbf{n}_i\| \sin(\sqrt{2}\gamma + \xi_i), & \sqrt{2}\gamma + \xi_i < \pi/2 \\ \|\mathbf{n}_i\|, & \sqrt{2}\gamma + \xi_i \geq \pi/2 \end{cases} \quad (5.63c)$$

where  $\xi_i \triangleq \arcsin(\tau / \|\mathbf{n}_i\|)$ ,  $\tau$  is the inlier threshold, and  $N'$  is the cardinality of the inlier set  $\mathcal{I}_1^*$ .

*Proof.* The completed proof is given in [Liu+22].  $\square$



**Algorithm 7:** BnB algorithm (solution to problem (5.60))

**Input:** Solution domain  $\mathbb{B}$ , inlier threshold  $\tau$ , candidate inlier correspondence set

$$\{(\hat{\mathbf{p}}_i, \hat{\mathbf{q}}_i)\}_{i=1}^{N'}$$
 of problem (5.48).

**Output:** Globally optimal solution  $\hat{\mathbf{C}}^*$ .

- 1 Generate a set of perpendicular bisector functions  $\{\mathbf{n}_i = [a_i, b_i, c_i]^T\}_{i=1}^{N'}$  by Eq. (5.58);
- 2 Initialize the queue of sub-branches  $q = \{\mathbb{B}\}$ ;
- 3 Initialize the lower bound  $L = 0$ , and the upper bound  $U = N'$ ;
- 4 Define function  $\delta(\cdot)$  returns the center of sub-branch;
- 5 **while**  $U - L > 0$  **do**
- 6     Select the sub-branch  $\mathbb{B}$  with the maximal upper bound from  $q$ ;
- 7     Divide  $\mathbb{B}$  into four sub-branches  $\{\mathbb{B}_j\}_{j=1}^4$ , and calculate each  $U(\mathbb{B}_j)$  and  $L(\mathbb{B}_j)$ ;
- 8     Insert the sub-branches with bounds into  $q$  and eliminate  $\mathbb{B}$  from  $q$ ;
- 9     Update  $U = \max\{U(\mathbb{B}_k)\}$  for all  $\mathbb{B}_k \in q$ ;
- 10    Update  $L = \max\{L, L(\mathbb{B}_k)\}$  for all  $\mathbb{B}_k \in q$ , if  $L(\mathbb{B}_k) > L$ , set  $\hat{\mathbf{C}}^* = \delta(\mathbb{B}_k)$ ;
- 11    Eliminate  $\mathbb{B}_k$  from  $q$  that  $U(\mathbb{B}_k) < L$ ;
- 12 **end**

Based on Proposition 5.3, the proposed BnB algorithm for the 2-DOF pole search sub-problem is given in Algorithm 7. To improve the algorithm efficiency, the input data of Algorithm 7 is set as the candidate inlier correspondence set of problem (5.48). After obtaining the optimal solution  $\hat{\mathbf{C}}^* = [c_x^*, c_y^*, w^*]^T$  by the proposed algorithm, we can transform it back to the 2D coordinate by

$$\mathbf{C}_0^* = \begin{bmatrix} \frac{c_x^*}{w^*} & \frac{c_y^*}{w^*} \end{bmatrix}^T \quad (5.64)$$

**Stage III: Voting for the Rotation Angle**

In the last stage, we aim to solve the remaining 1-DOF rotation angle estimation sub-problem. According to Eq. (5.57), we can observe that the rotation angle is the angle between two vectors  $(\hat{\mathbf{p}}_i - \mathbf{C}_0)$  and  $(\hat{\mathbf{q}}_i - \mathbf{C}_0)$  in 2D plane  $\Omega_0$ . Given optimal pole  $\mathbf{C}_0^*$ , the rotation angle for the  $i$ -th candidate inlier correspondence can be computed as

$$\theta_i = \angle(\hat{\mathbf{p}}_i - \mathbf{C}_0^*, \hat{\mathbf{q}}_i - \mathbf{C}_0^*). \quad (5.65)$$

Inspired by the *uniform grid approach* [Nes+18], we take  $s$  equally spaced 1D grids on  $[0, 2\pi]$  whose centers are  $\{\theta_k = (2k-1)\pi/s\}_{k=1}^s$  as the complete solution domain of rotation angle. Therefore, given the set of candidate inlier correspondences from problem (5.60), the third sub-problem aims at estimating the optimal  $\theta$  to maximize the cardinality of the inlier set, which is defined by

$$\begin{aligned} & \max_{\theta \in \{\theta_k\}_{k=1}^s, \mathcal{I}_3 \subseteq \mathcal{I}_2^*} |\mathcal{I}_3| \\ & \text{s.t.} \quad \left| \theta - \angle(\hat{\mathbf{p}}_i - \mathbf{C}_0^*, \hat{\mathbf{q}}_i - \mathbf{C}_0^*) \right| \leq \zeta, \quad \forall i \in \mathcal{I}_3 \end{aligned} \quad (5.66)$$

where  $\mathcal{I}_3$  is the inlier set extracted from the cardinality-maximized inlier set  $\mathcal{I}_2^*$ , and  $\zeta \triangleq \pi/s$  is the inlier threshold. We can apply *voting-based method* [Gle+14; Sca11; YSC20; Yan+23] to efficiently address this consensus maximization problem. For  $N'' \triangleq |\mathcal{I}_2^*|$  candidate inlier correspondences, we can vote  $\theta^*$  with the largest consensus set  $\mathcal{I}_3^*$  from  $\{\theta_i\}_{i=1}^{N''}$ . Due to its logical simplicity, we omit the global voting algorithm here. Considering the trade-off between accuracy and efficiency, we set  $s = 360$  as an invariant parameter. A visualized illustration of the proposed global voting method is presented in Fig. 5.7.

After obtaining optimal  $\theta^*$ , according to Eq. (5.56) and Eq. (5.57), we can derive that

$$\hat{\mathbf{t}}_{\perp}^* = (\mathbf{I} - \hat{\mathbf{R}}_{\theta}^*)\mathbf{C}_0^* \quad (5.67)$$

where  $\mathbf{I}$  is an identity matrix. Then, according to Eq. (5.54) and Eq. (5.55), we have

$$\mathbf{R}_{\theta}^* = \begin{bmatrix} \cos \theta^* & -\sin \theta^* & 0 \\ \sin \theta^* & \cos \theta^* & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \tilde{\mathbf{t}}_{\perp}^* = \begin{bmatrix} (\mathbf{I} - \hat{\mathbf{R}}_{\theta}^*)\mathbf{C}_0^* \\ 0 \end{bmatrix} \quad (5.68)$$

Finally, according to Eq. (5.42) and Eq. (5.52), we can obtain the optimal rotation and translation by

$$\mathbf{R}^* = (\mathbf{R}_{v_q}^{e_z})^T \mathbf{R}_{\theta}^* \mathbf{R}_{v_q}^{e_z} \mathbf{R}_{v_p}^{v_q} \quad (5.69)$$

$$\mathbf{t}^* = \mathbf{t}_{\perp}^* + \mathbf{t}_{\parallel}^* = (\mathbf{R}_{v_q}^{e_z})^T \tilde{\mathbf{t}}_{\perp}^* + \mathbf{t}_{\parallel}^* \quad (5.70)$$

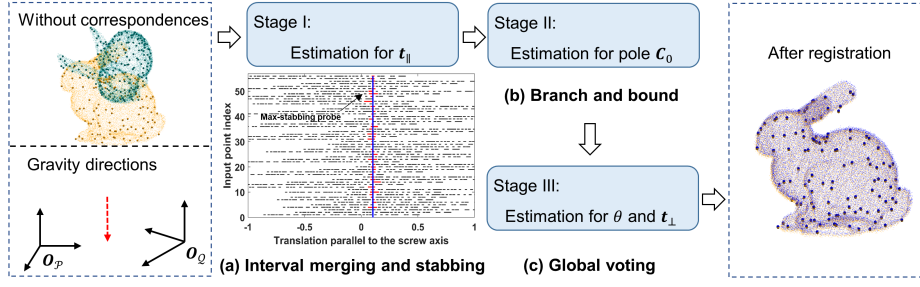
### 5.3.3 Simultaneous Pose and Correspondence Registration

Since the performance limitation of current 3D feature matching methods, we have to face the situation that the correspondences are unknown in some practical applications, such as when the point clouds are not sampled densely from smooth surfaces [Yan+16]. Therefore, we explore the feasibility and potential of the proposed approach in solving the more challenging SPCR problem in this section.

We assume that point clouds  $\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^M$  and  $\mathcal{Q} = \{\mathbf{q}_j\}_{j=1}^N$  are the source and target point cloud, respectively. Following [Bus+16; Cam+18], and assuming known gravity directions, we have the following consensus maximization problem

$$\begin{aligned} \max_{\theta, \mathbf{C}, \mathbf{t}_{\parallel}, \mathcal{I} \subseteq \mathcal{S}} & |\mathcal{I}| \\ \text{s.t.} & \exists j \in \mathcal{K}, \|\mathbf{S}(\mathbf{p}_i') + \mathbf{t}_{\parallel} - \mathbf{q}_j\| \leq \epsilon, \forall i \in \mathcal{I}, \end{aligned} \quad (5.71)$$

where  $\mathcal{S} = \{1, \dots, M\}$  and  $\mathcal{K} = \{1, \dots, N\}$  are the sets of indices for point clouds  $\mathcal{P}$  and  $\mathcal{Q}$ , respectively. According to the derivation in Section 5.3.2, we can obtain a new consensus maximization sub-problem for estimating the translation parallel to the screw axis without



**Figure 5.11:** Calculation flow of the proposed three-stage method for correspondence-free registration. The bolded points in the point cloud represent the points after downsampling. For the interval merging and stabbing part, the merged intervals are represented by the black line segments, while the intervals crossed by the max-stabbing probe are depicted as the red line segments.

correspondences, which is defined by

$$\begin{aligned}
 & \max_{l, \mathcal{I}_1 \subseteq \mathcal{S}} |\mathcal{I}_1| \\
 & \text{s.t. } \exists j \in \mathcal{K}, l \in [l_{ij}^-, l_{ij}^+], \forall i \in \mathcal{I}_1, \\
 & \quad l_{ij}^- = -\delta - \mathbf{v}_q^T (\mathbf{p}'_i - \mathbf{q}_j), \\
 & \quad l_{ij}^+ = \delta - \mathbf{v}_q^T (\mathbf{p}'_i - \mathbf{q}_j).
 \end{aligned} \tag{5.72}$$

For sub-problem (5.72), we introduce an *interval merging* algorithm [De +97] (Algorithm 8) to effectively reduce the number of input intervals before interval stabbing. This operation can avoid the all-to-all correspondence assumption [YSC20] in which the assumed correspondence-based registration problem is extremely outlier-contaminated (with high input number and high outlier rate), thereby affecting the efficiency. After interval merging, the max-stabbing probe can only cross through at most one interval for each point  $\mathbf{p}'_i$ . From the perspective of inlier set cardinality, each point  $\mathbf{p}'_i$  contributes a maximum of 1 to the cardinality of the inlier set. Therefore, the interval merging algorithm is executed for each point  $\mathbf{p}'_i$ , followed by executing the interval stabbing algorithm (Algorithm 6) for all merged intervals. The subplot of Fig. 5.11 illustrates an example of the visualization results about interval merging and stabbing.

After solving the correspondence-free sub-problem (5.72), we can obtain candidate inlier correspondences that only satisfy the 1-DOF constraint. While it may not be possible to identify outliers that coincidentally satisfy this constraint, we can readily employ the proposed correspondence-based methods to tackle the second and third sub-problems (Eq. (5.60) and Eq. (5.66)), enabling the estimation of the remaining translation and the rotation angle. The three-stage calculation flow for the SPCR problem is shown in Fig. 5.11.

## 5.4 LiDAR Registration Experiments

**Algorithm 8:** Interval merging algorithm

---

**Input:** Intervals  $\left\{ \left\{ \left[ l_{ij}^-, l_{ij}^+ \right] \right\}_{i=1}^M \right\}_{j=1}^N$

**Output:** Merged intervals  $\mathcal{L} = \{[\mu_k, \nu_k]\}_{k=1}^K$ , where  $K = \sum_{i=1}^M N_i$

- 1 Initialize the index  $i = 1$ ;
- 2 Initialize the list of merged intervals  $\mathcal{L} = \emptyset$ ;
- 3 **while**  $i \leq M$  **do**
- 4     Sort the intervals  $\left\{ \left[ l_{ij}^-, l_{ij}^+ \right] \right\}_{j=1}^N$  by left side  $l_{ij}^-$  in ascending order;
- 5     Initialize  $\mu_1 = l_{i1}^-$ ,  $\nu_1 = l_{i1}^+$ ;
- 6     Initialize the index  $j = 1$ ,  $k = 1$ ;
- 7     **while**  $j < N$  **do**
- 8         **if**  $l_{ij}^+ \geq l_{i(j+1)}^-$  **then**
- 9              $\nu_k = \max(l_{ij}^+, l_{i(j+1)}^+)$ ;
- 10         **else**
- 11              $k = k + 1$ ;
- 12              $\mu_k = l_{i(j+1)}^-$ ,  $\nu_k = l_{i(j+1)}^+$ ;
- 13         **end**
- 14          $j = j + 1$ ;
- 15     **end**
- 16     Add the merged intervals into list  $\mathcal{L}$ , i.e.,  $\mathcal{L} = \mathcal{L} \cup \{[\mu_k, \nu_k]\}_{k=1}^{N_i}$ , where  $N_i$  is the number of merged intervals for  $i$ ;
- 17 **end**

---

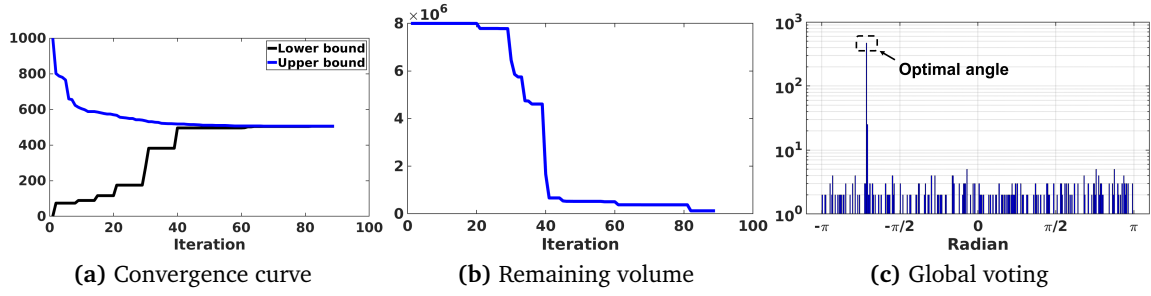
**5.4.1 Experiments for the Two-Stage Method**

This section compares the proposed two-stage method with several existing methods for the correspondence-based registration problem on both synthetic and real-world data. The extended method is compared with several benchmark methods regarding the SPCR problem on synthetic data. The proposed method is implemented in C. All experiments are performed on a laptop with an i7-9750H CPU and 16GB RAM.

**Experimental Setting**

All the methods compared are listed below:

- BnB [Cai+19]: A joint 4-DOF BnB-based method implemented in C++ for correspondence-based registration.
- RANSAC-2pt [FB81]: A 4-DOF version of RANSAC (using minimal 2-points) implemented in C++ for correspondence-based registration. The confidence level is 0.99, and the maximum number of iterations is  $1e7$ .
- FGR-4DOF [ZPK16]: A 4-DOF version of FGR implemented in C++ for correspondence-based registration. The annealing rate is 1.1.



**Figure 5.12:** Efficient translation search by the proposed BnB and global voting for rotation estimation.

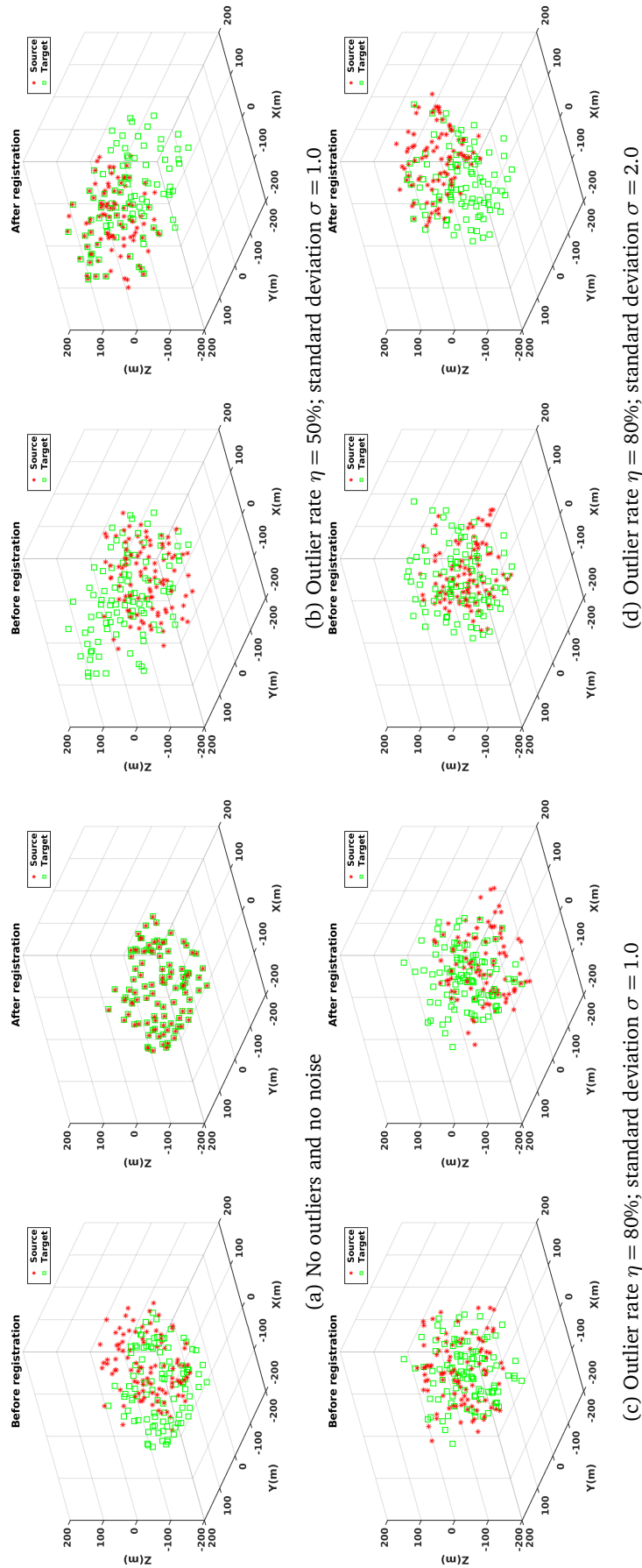
- GTR [ART10]: A stochastic outlier removal method implemented in C++ for correspondence-based registration. The maximum number of iterations is 10000.
- K4PCS [TWS14]: A 4PCS-based variant based on keypoints implemented in C++ for the SPCR problem. The score threshold is 0.001.
- GoICP [Yan+16]: A joint 6-DOF nested BnB-based method implemented in C++ for the SPCR problem. Both source and target point clouds have to be normalized to fit in  $[-1, 1]^3$ , such that the translation domain is  $[-0.5, 0.5]^3$ . The rotation domain is  $[-\pi, \pi]^3$ , and the mean squared error (MSE) convergence threshold is 0.001.
- GoICPT [Yan+16]: A variant of 6-DOF GoICP with trimming (trimming fraction 10%).
- ICP [BM92]: A classic EM-type method implemented in the *pcregistericp* function of MATLAB for the SPCR problem. The translation and rotation tolerance between consecutive iterations is  $[0.01, 0.05]$ , and the maximum number of iterations is 100.
- CPD [MS10]: A robust EM-type method implemented in C for the SPCR problem. The weight of noise and outliers is 0.1, and the tolerance stopping criterion is  $1e-8$ .
- GMMReg [JV10]: A robust EM-type method implemented in C for the SPCR problem. The maximum number of iterations is 1000.

Among the correspondence-based registration methods, BnB and FGR-4DOF are deterministic. Conversely, RANSAC-2pt and GTR are non-deterministic due to sampling uncertainty. The performance and accuracy of each algorithm are evaluated by 1) rotation error  $e_{\text{rot}}$ , 2) translation error  $e_{\text{trans}}$ , and 3) success rate  $r_{\text{thr}}$ . The definitions are as follows:

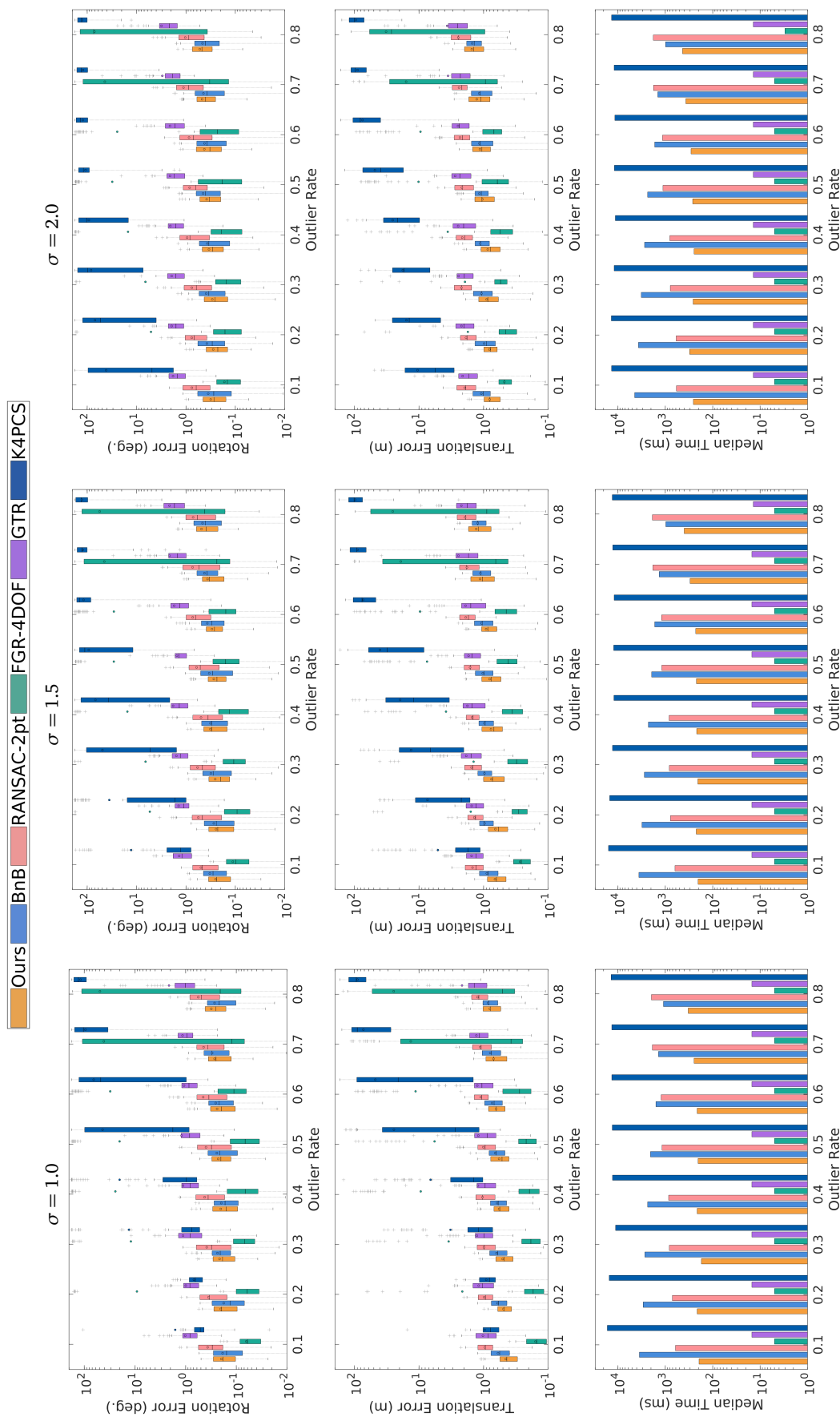
$$e_{\text{rot}} = \arccos\left(\frac{\text{Tr}(\mathbf{R}_{\text{gt}}^{-1}\mathbf{R}^*) - 1}{2}\right), \quad (5.73)$$

$$e_{\text{trans}} = \|\mathbf{t}_{\text{gt}} - \mathbf{t}^*\|, \quad (5.74)$$

$$r_{\text{thr}} = \frac{T^+}{T_{\text{total}}}, \quad (5.75)$$



**Figure 5.13:** Visualization of the synthetic point cloud data. Each subplot (a)-(d) contains two figures (before and after registration with our method) and has different parameter settings, as indicated in the subheadings. The source point cloud is shown in red, and the target point cloud is shown in green. The rotation error  $e_{rot}(\circ)$  and translation error  $e_{trans}(m)$  for each example are: (a) 0.024 and 0.144, (b) 0.228 and 0.787, (c) 0.323 and 0.953, (d) 0.504 and 1.326, respectively. It is evident that the rotation and translation errors increase significantly as the outlier rate and noise level increase.



**Figure 5.14:** Controlled experiments with normal ( $\sigma = 1.0$ ) and high ( $\sigma = 1.5, 2.0$ ) noise levels; in each group, the outlier rate is  $\eta = \{0.1, 0.2, \dots, 0.8\}$ . The results include rotation error, translation error, and median time.

**Table 5.1:** Success rates  $r_1$  and  $r_2$  (%) with successful cases satisfying Threshold 1 ( $e_{\text{rot}} \leq 0.5^\circ, e_{\text{trans}} \leq 0.2\text{m}$ ) and Threshold 2 ( $e_{\text{rot}} \leq 5^\circ, e_{\text{trans}} \leq 2\text{m}$ ), respectively, on the ETH dataset

Dataset	Ours		BnB		RANSAC-2pt		FGR-4DOF		GTR		K4PCS	
	$r_1$	$r_2$	$r_1$	$r_2$	$r_1$	$r_2$	$r_1$	$r_2$	$r_1$	$r_2$	$r_1$	$r_2$
<i>Arch</i>	<b>100</b>	<b>100</b>	80	80	80	<b>100</b>	40	40	0	0	0	60
<i>Courtyard</i>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	0	43	57	71
<i>Facade</i>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	33	83
<i>Office</i>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	20	40	40	60	60	80	0	80
<i>Trees</i>	<b>100</b>	<b>100</b>	83	83	83	<b>100</b>	33	33	17	33	17	67

where  $\mathbf{t}_{\text{gt}}$  and  $\mathbf{R}_{\text{gt}}$  are the ground truth,  $\mathbf{t}^*$  and  $\mathbf{R}^*$  are the estimated solutions,  $Tr(\cdot)$  is the trace of a matrix,  $T_{\text{total}}$  is the total number of trials and  $T^+$  is the number of successful cases satisfying the predefined threshold.

For a simple demonstration of the convergence of the proposed two-stage method, a pair of synthetic point clouds ( $N = 1000$ ) is used as the input, and the figures about the convergence curve, remaining volume, and global voting are plotted in Fig. 5.12. It is evident that the gap between the lower and upper bounds is converging to zero, and the proposed method converges to the optimal solution after dozens of iterations. The remaining volume of the translation domain decreases rapidly until convergence. Furthermore, the proposed global voting method can successfully search the optimal angle  $\theta^*$  from the interval  $[-\pi, \pi]$ .

### Synthetic Data Experiments

In this section, the performance of the proposed two-stage method is evaluated in both correspondence-based registration and SPCR problems by conducting synthetic data experiments. The advantage of synthetic data experiments is that the precise ground truth is available, which means that detailed evaluations can be made and theoretical performance can be achieved.

**Data generation and settings.** Initially, the source point cloud is generated by creating  $N$  random points distributed in the cube  $[-100, 100]^3$ . Random rotation in the interval  $[-\pi, \pi]$  and random translation in the cube  $[-100, 100]^3$  are applied to the source point cloud to obtain the corresponding target point cloud. Then, partial points in both source and target point clouds are replaced with arbitrarily generated points in the cube  $[-100, 100]^3$  to simulate outliers, and the outlier rate is defined as  $\eta$ . Finally, the zero-mean Gaussian noise with standard deviation  $\sigma$  (noise level) is added to both source and target point clouds. The visualization of the synthetic point cloud ( $N = 100$ ) before and after registration using the proposed two-stage method is shown in Fig. 5.13. In addition, the translation search domain of the proposed method is  $[-200, 200]^3$  in all synthetic data experiments. The gravity directions in the proposed method are prior known. The settings of each algorithm are given in Section 5.4.1. The distance-based inlier threshold for all compared methods is equal to the standard deviation  $\sigma$  of noise. Similarly, the angle-based inlier threshold for the proposed method is also selected according to the standard deviation of the noise.



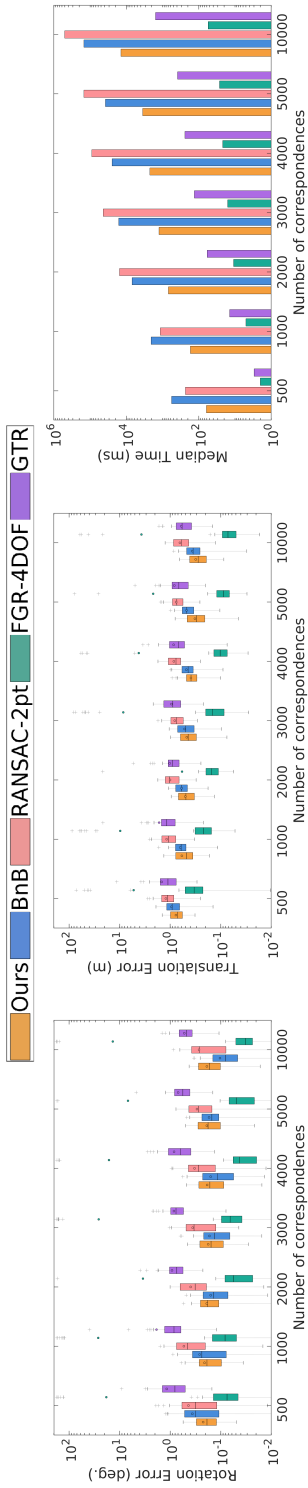


Figure 5.15: Controlled experiments with different numbers of correspondences ( $N = \{500, 1000, \dots, 10000\}$ ). The outlier rate is  $\eta = 0.5$  and the noise level is  $\sigma = 1.0$ .

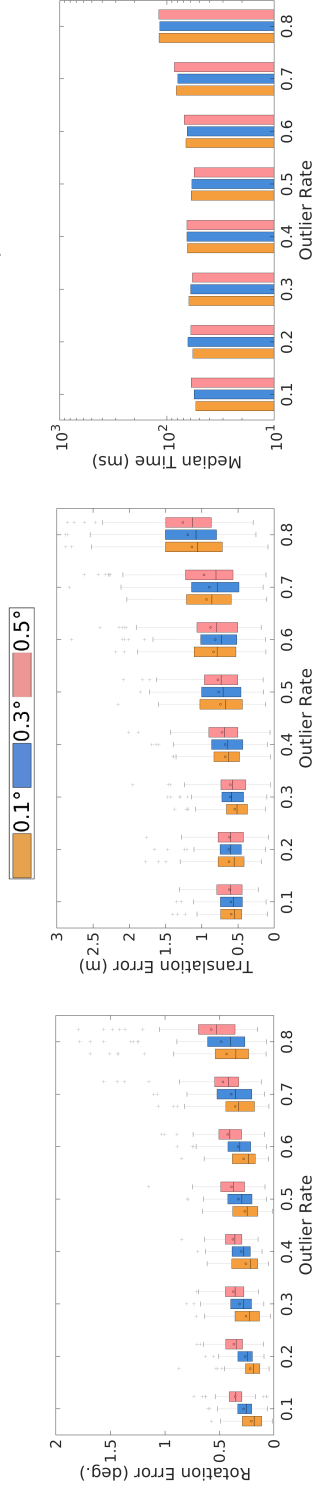


Figure 5.16: Controlled experiments with different biased angles ( $0.1^\circ, 0.3^\circ, 0.5^\circ$ ) in gravity directions and different outlier rates  $\eta = \{0.1, 0.2, \dots, 0.8\}$ .

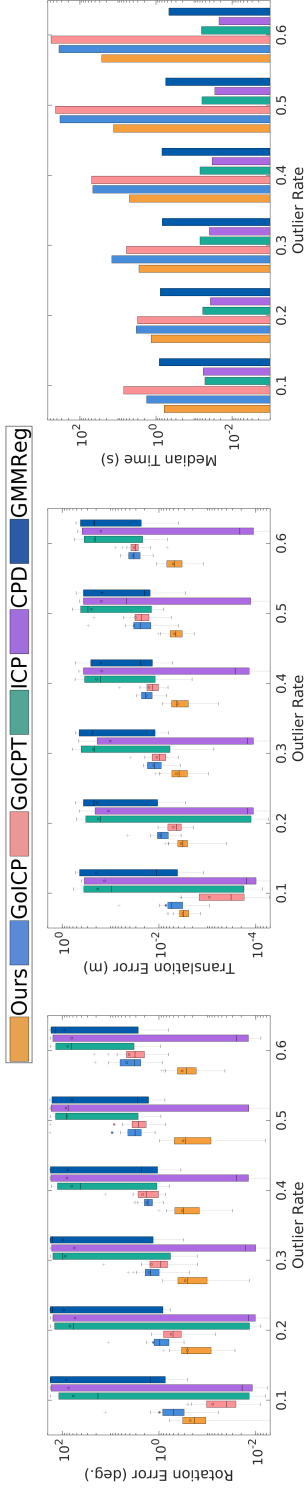


Figure 5.17: Challenging SPCR experiments with different outlier rates  $\eta = \{0.1, 0.2, \dots, 0.6\}$  in the Bunny dataset.

**Table 5.2:** The left part is detailed information about the ETH dataset. The right part is the registration running time (ms) for each pair of scans on the ETH dataset.

Dataset	Number of points ( $10^6$ )	Number of keypoints	Number of correspondences	FMP	Ours	BnB	RANSAC-2pt	FGR-4DOF	GTR	K4PCS	
<i>Arch</i>	s1-s2	23.56-30.90	7905-4783	19860	6793	2.6	635	9582	115	5200	553869
	s2-s3	30.90-25.25	4783-7146	19328	5004	9.7	280	46081	98	4815	89452
	s3-s4	25.25-29.45	7146-5337	22300	7534	21.6	616	6445	121	6153	245060
	s4-s5	29.45-27.96	5337-4676	15508	3809	61.9	786	43127	82	3249	129117
	s1-s5	23.56-27.96	7905-4676	17999	5894	5.6	1238	95211	100	4195	365587
<i>Courtyard</i>	s1-s2	13.32-18.80	3944-7451	28640	9109	27.0	860	477	171	9175	47901
	s2-s3	18.80-12.71	7451-3316	24208	7132	22.1	2737	436	127	6826	418675
	s3-s4	12.71-12.15	3316-4083	20047	4059	158.0	3893	121	106	5042	49795
	s4-s5	12.15-16.75	4083-1400	11012	1670	43.0	1026	110	56	2098	159567
	s5-s6	16.75-11.99	1400-3756	10640	1385	20.9	581	166	60	1712	3589
	s6-s7	11.99-11.17	3756-4275	21979	4527	58.8	1145	310	125	6243	70327
	s7-s8	11.17-11.45	4275-5609	26852	8904	14.2	1535	300	144	8565	70002
<i>Facade</i>	s1-s2	25.08-15.25	651-1066	3253	177	0.5	18	143	16	163	2859
	s2-s3	15.25-15.79	1066-797	3709	166	0.9	10	118	24	201	11471
	s3-s4	15.79-23.48	797-672	2965	101	2.1	22	78	16	135	4840
	s4-s5	23.48-16.74	672-629	2872	93	0.8	34	47	14	122	3834
	s5-s6	16.74-21.46	629-674	2823	92	0.8	46	57	15	117	3919
	s6-s7	21.46-20.79	674-506	2499	78	1.2	13	148	12	89	3051
<i>Office</i>	s1-s2	10.72-10.71	482-434	2092	44	1.2	12	73	12	84	3835
	s2-s3	10.71-10.73	434-470	2053	51	0.8	23	22	8	62	5998
	s3-s4	10.73-10.69	470-411	1797	45	0.5	9	109	9	53	4013
	s4-s5	10.69-10.75	411-468	1882	60	1.2	25	204	7	77	12971
	s1-s5	10.72-10.75	482-468	2286	72	3.9	39	190	20	91	11303
<i>Trees</i>	s1-s2	19.63-19.60	11121-11324	33485	19172	0.3	793	151770	221	2578	983485
	s2-s3	19.60-19.77	11324-11287	33740	19635	4.8	507	15816	217	2096	21669
	s3-s4	19.77-20.39	11287-10178	31995	18809	3.8	207	320304	208	2065	1000000
	s4-s5	20.39-20.48	10178-10982	32878	16915	15.2	617	16350	224	1911	757556
	s5-s6	20.48-21.61	10982-9274	30898	17299	5.5	480	22998	183	1644	906439
	s1-s5	19.63-20.48	11121-10982	32838	18774	130.4	1349	579404	187	1944	986573

**Controlled experiments with different outlier rates and noise levels.** In this section, the robustness and efficiency experiments are conducted for Ours, BnB, RANSAC-2pt, FGR-4DOF, GTR, and K4PCS at different outlier rates and noise levels. We first randomly generate  $N = 1000$  points for both the source and target point clouds to simulate the set of candidate correspondences. We then set up three groups of experiments with different noise levels, including normal ( $\sigma = 1.0$ ) and high ( $\sigma = 1.5, 2.0$ ). In each group, the outlier rates are sequential, i.e.,  $\eta = \{0.1, 0.2, \dots, 0.8\}$ . To observe the theoretical performance, the experiments are repeated 100 times for each method under different conditions. The rotation errors and translation errors of all methods are compared in the box plot, as shown in the first two rows of Fig. 5.14, and the median running times of all methods are compared in the bar chart, as shown in the last row of Fig. 5.14. The gray '+' in the box plot denotes the outlier in 100 trials.

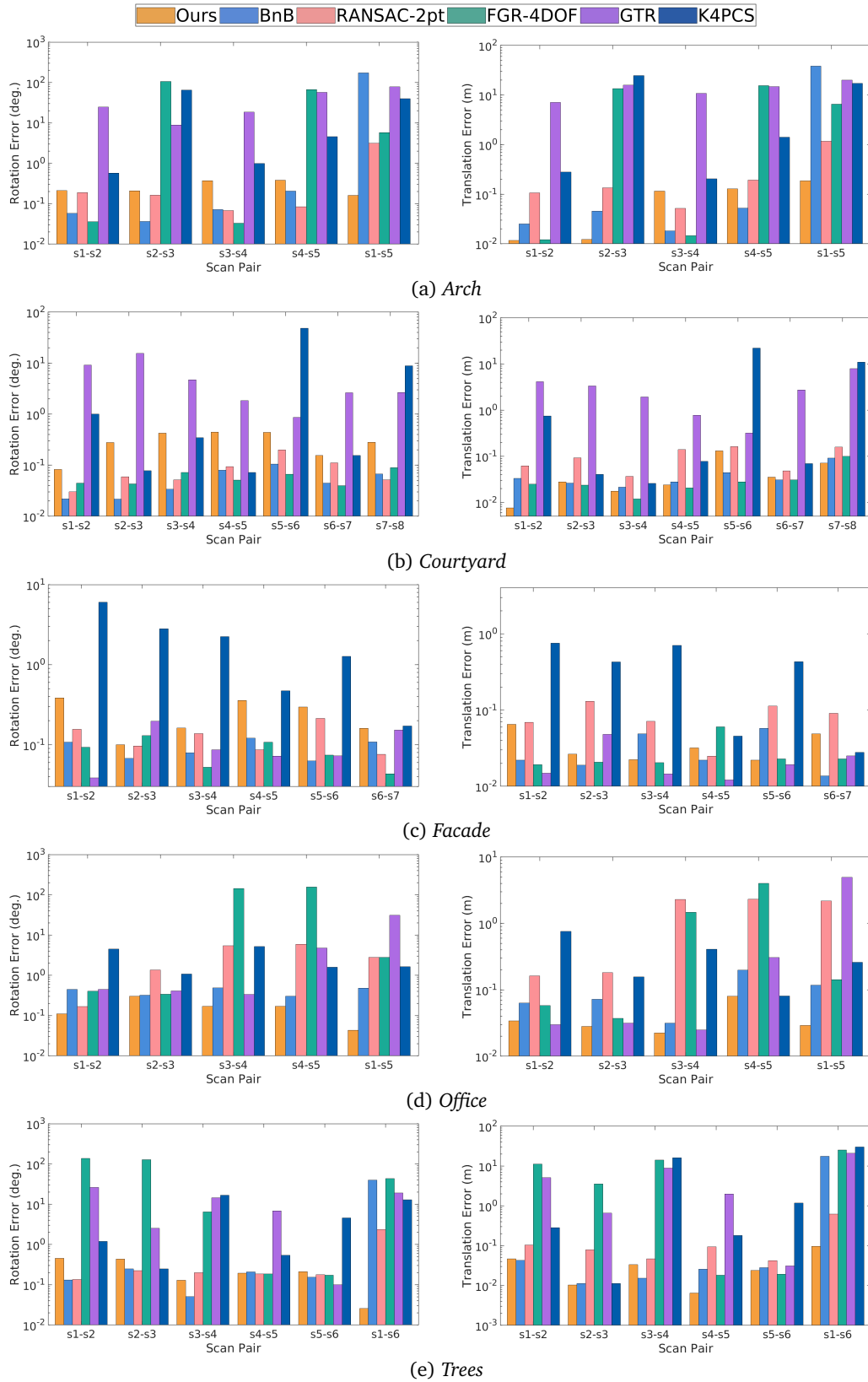
A comparison of the rotation error and translation error shows that Ours and BnB perform well under all outlier rates. Moreover, in most cases, the average errors of Ours are smaller than those of all methods except FGR-4DOF. However, deterministic FGR-4DOF is prone to

delivering erroneous results (local optimum) in some cases, especially when the outlier rate exceeds 70%, resulting in very large average errors. In contrast, the outlier results in 100 trials for each set of experiments illustrate that the non-deterministic RANSAC-2pt occasionally produces unsatisfactory results with significant errors. This is probably because of the sampling uncertainty of RANSAC. Similar to RANSAC-2pt, stochastic GTR only generates a satisfactory solution with a certain probability. Also, K4PCS, which solves the SPCR problem, does not perform stably, even with a low outlier rate, and its average errors are the largest of all the methods. On the other hand, our decomposed method is about 2 to 18 times faster than the joint 4-DOF BnB in the comparison of median run times, which is almost an order of magnitude faster. Thanks to the reduced search domain, the proposed method displays significant efficiency. Although FGR-4DOF and GTR are the two fastest methods, they return unsatisfactory solutions with a certain probability. In general, our decomposition-based method is more efficient than the approaches with comparable robustness to ours.

**Controlled experiments with different numbers of correspondences.** In this section, the accuracy and efficiency experiments are conducted for Ours, BnB, RANSAC-2pt, FGR-4DOF, and GTR with different numbers of correspondences. The set of candidate correspondences is generated sequentially with the number  $N = \{500, 1000, \dots, 10000\}$ . Then, the outlier rate is set to  $\eta = 0.5$ , and the noise level is set to  $\sigma = 1.0$ . The experiments are repeated 50 times for each method under different conditions. Then, the rotation errors and translation errors in the box plot are compared, and the median running times are compared in the bar chart, as shown in Fig. 5.15. Please note that the results of K4PCS are not plotted because the duration is over 1000 seconds in most experiments.

For the deterministic methods, it can be observed that the accuracy of Ours is higher than BnB and lower than FGR-4DOF. However, the average rotation and translation errors of FGR-4DOF are the highest because FGR-4DOF occasionally leads to registration failure. The non-deterministic methods RANSAC-2pt and GTR also generate correct solutions in most cases, but their accuracy is lower than the deterministic methods. The median time plot indicates that the proposed method is about 9 to 10 times faster than joint 4-DOF BnB under different correspondence numbers. Moreover, as the number of correspondences increases, the running time of Ours increases more slowly than RANSAC-2pt, which means that, for instance, RANSAC-2pt is similar to Ours at 500 correspondences but slowest at 10000 correspondences. These experiment results illustrate the superiority of the proposed two-stage method in terms of accuracy and efficiency.

**Robustness to gravity direction biases.** In this section, biased gravity directions are generated to simulate the measurement biases of IMUs. Experiments are then conducted to verify the robustness of the proposed two-stage method against gravity direction biases. Specifically, we first randomly generate  $N = 500$  correspondences and set up one group of experiments with noise level  $\sigma = 1.0$ . The outlier rates in the experiments are sequential, i.e.,  $\eta = \{0.1, 0.2, \dots, 0.8\}$ . The biased angles are set as  $0.1^\circ, 0.3^\circ$ , and  $0.5^\circ$ , respectively. Please note that the biased angle is randomly generated in an arbitrary direction. The experiment



**Figure 5.18:** Rotation error and translation error of all registration methods on the ETH dataset. Subfigures (a)-(e) show the registration results of each scan pair for *Arch*, *Courtyard*, *Facade*, *Office*, and *Trees*.

is repeated 100 times for each experimental setting. In this experiment, only the proposed method with different biased angles is compared, since the other methods do not rely on the gravity direction. Then, the rotation and translation errors are recorded in the box plot, and the median running times are recorded in the bar chart, as shown in Fig. 5.16. As can be seen, different biased angles have a certain effect on the rotation error but have little impact on the translation error. In addition, the median time of the proposed method is insensitive to the biased angle in gravity directions. Overall, these experiments prove that the proposed two-stage method is robust to the biased angle in gravity directions and remains robust to outliers when gravity directions are biased.

**Challenging SPCR experiments.** In this section, the challenging SPCR experiments are conducted for Ours, GoICP, GoICPT, ICP, CPD, and GMMReg at different outlier rates. The Bunny dataset from the Stanford 3-D Scanning Repository [CL96] is used for this experiment. The Bunny dataset is first downsampled to 234 points as the source point cloud. Then, a random transformation is applied to the source point cloud to obtain the target point cloud. In accordance with the requirement of GoICP, the source and target point clouds are normalized to fit in  $[-1, 1]^3$ . Next, a fraction of the points in the source point cloud are removed to simulate the outlier, with the remaining points as inliers. The zero-mean Gaussian noise with  $\sigma = 0.001$  is also added to both source and target point clouds. The experiments are repeated 30 times at each outlier rate, i.e.,  $\eta = \{0.1, 0.2, \dots, 0.6\}$ .

The rotation and translation errors are recorded in the box plot, and the median running times are recorded in the bar chart, as shown in Fig. 5.17. It can be observed that ICP and GMMReg converge to the incorrect solution in most cases, leading to large average errors. CPD is better than ICP and GMMReg but is still prone to incorrect transformation in many cases. GoICP and GoICPT are global methods without initialization, and they perform much better than ICP, CPD, and GMMReg. However, in most cases, the proposed method achieves the fewest average errors. It is worth noting that GoICPT is a variant of GoICP with prior trimming fraction (10%), such that its errors are the lowest at 10% outlier rate. Moreover, the proposed method is faster than GoICP and GoICPT because our search domain for BnB is 3-DOF, while theirs is 6-DOF. This experiment shows that 1) the proposed method can also solve the challenging SPCR problem, 2) it is more robust than the existing local methods, e.g., ICP, CPD, and GMMReg, and 3) it is more accurate than the existing global methods, e.g., GoICP.

### Real-World Data Experiments

In this section, the performance of the proposed two-stage method is verified on the ETH dataset [TS+12], a large-scale terrestrial LiDAR dataset. It incorporates several indoor and outdoor sub-datasets, including *Arch*, *Courtyard*, *Facade*, *Office*, and *Trees*. Each sub-dataset contains several different point clouds measured from different positions, and the experiment registers pairwise point clouds in each sub-dataset. The overlapping ratios of *Arch*, *Courtyard*, *Facade*, *Office*, and *Trees* are 30 – 40%, 40 – 70%, 60 – 70%, > 80%,  $\approx 50\%$ , respectively [LHA20]. A similar preparation strategy to that of [Cai+19] is utilized. Specifically, ISS

keypoints [Zho09] and FPFH descriptors [RBB09] are first used to generate the putative correspondences. The number of points, keypoints, and correspondences for each scan pair in the ETH dataset are shown in the left part of Table 5.2. Then, Ours, BnB, RANSAC-2pt, FGR-4DOF, GTR, and K4PCS are used to register pairwise point clouds from each sub-datasets. In the process of operating Ours and BnB, FMP [Cai+19] is used to accelerate the convergence of BnB-based methods. Please note that the Z-axis of each scan is aligned by the tilt compensator in the terrestrial LiDAR scanner [Cai+19], thus the gravity direction employed in the proposed method is set as  $[0, 0, -1]^T$ . The ground truth is contained in the dataset. The inlier thresholds are set to 0.1m and  $0.5^\circ$  for distance-based and angle-based methods, respectively.

Fig. 5.18 shows the rotation and translation errors for different scan pairs in each sub-dataset. In most cases, the errors of Ours are not the lowest since only rough gravity directions are employed in Ours. However, viewed overall, the errors of Ours are acceptable in practice in all cases. On the other hand, the results of Ours can also be refined with EM-type methods (e.g., ICP) to reduce errors. We set Threshold 1 ( $e_{\text{rot}} \leq 0.5^\circ, e_{\text{trans}} \leq 0.2\text{m}$ ) and Threshold 2 ( $e_{\text{rot}} \leq 5^\circ, e_{\text{trans}} \leq 2\text{m}$ ) to decide the success rate, as shown in Table 5.1. The success rates with two different thresholds of Ours always stay at 100%. The second best method is deterministic BnB, which maintains a success rate of over 80%. FGR-4DOF, another deterministic method, and all non-deterministic methods such as RANSAC-2pt, GTR, and K4PCS randomly generate incorrect results. Table 5.2 shows the running time of each method. Compared with BnB, the running time of Ours is about 10 to 250 times faster, which indicates the high efficiency of the proposed two-stage method.

## 5.4.2 Experiments for the Three-Stage Method

### Experimental Setting

In this section, we compare the performance of the proposed three-stage method with SOTA correspondence-based approaches by utilizing both synthetic and real-world data. Besides, we compare our extended SPCR method with different correspondence-free approaches. All experiments are implemented in a PC with an AMD 5600x CPU and 32GB RAM.

**Compared methods.** In the correspondences-based registration experiments, our method is compared with the 4-DOF as well as 6-DOF registration methods, including:

- FMP+BnB [Cai+19]: A joint 4-DOF BnB method using fast match pruning (FMP) as the preprocessing step, programmed in C++.
- BnB [Cai+19]: A joint 4-DOF BnB method without fast match pruning.
- RANSAC [FB81]: A typical consensus maximization approach, which is customized for 4-DOF registration and programmed in C++. The maximum number of iterations is set to  $10^7$ .
- FGR [ZPK16]: A fast M-estimation method, which is customized for 4-DOF registration and programmed in C++. The annealing rate is set to 1.1.

- GTA [ART10]: A 6-DOF outlier removal method based on the game-theoretic framework, programmed in C++.
- GROR [Yan+22a]: A 6-DOF fast outlier removal method based on the reliability of the correspondence graph, programmed in C++.

In addition, our extended method is compared with the following global and local methods for the SPCR problem:

- GO-ICP [Yan+16]: A globally optimal approach that combines global BnB and local ICP, programmed in C++.
- GO-ICPT [Yan+16]: A variant of GO-ICP with a specified trimming percentage for outlier removal.
- ICP [BM92]: A classic rigid registration algorithm that finds the optimal transformation by solving the least squares problem during each iteration, implemented in MATLAB Toolbox.
- CPD [MS10]: A probabilistic method that transfers the registration as a probability density estimation problem, programmed in C language.
- Gmmreg [JV10]: A probabilistic registration method that represents the input point clouds as GMMs and aligns them, programmed in C language.

In both correspondence-based and SPCR experiments, the proposed three-stage method is consistently referred to as **Ours**, which is implemented in Matlab 2022b.

**Evaluation metrics.** Following [Cai+19; Yan+22a], we employ running time  $T$ , rotation error  $RE$ , translation error  $TE$ , and success rate  $SR$  to evaluate the registration performance. The calculations of  $RE$  and  $TE$  are shown as below:

$$RE = \arccos\left(\frac{\text{Tr}(\mathbf{R}_{gt}^T \mathbf{R}^*) - 1}{2}\right) \quad (5.76)$$

$$TE = \|\mathbf{t}_{gt} - \mathbf{t}^*\| \quad (5.77)$$

where  $\mathbf{R}_{gt}$  and  $\mathbf{t}_{gt}$  are ground truth rotation and translation,  $\mathbf{R}^*$  and  $\mathbf{t}^*$  are estimated results, and  $\text{Tr}(\cdot)$  is the trace of a matrix. The point clouds are successfully aligned when  $RE$  and  $TE$  are within the predefined thresholds.

### Synthetic Data Experiments

To verify the theoretical performance of the proposed three-stage method, we first conduct experiments using synthetic data. The source point cloud is randomly generated by creating  $N$  points distributed in the cube  $[-1, 1]^3$ . Then, the source point cloud is transformed by a random rotation, whose angle is within  $[-\pi, \pi]$ , and a random translation in  $[-1, 1]^3$  to generate the target point cloud. Due to the limitations of [Cai+19], the rotation axis is

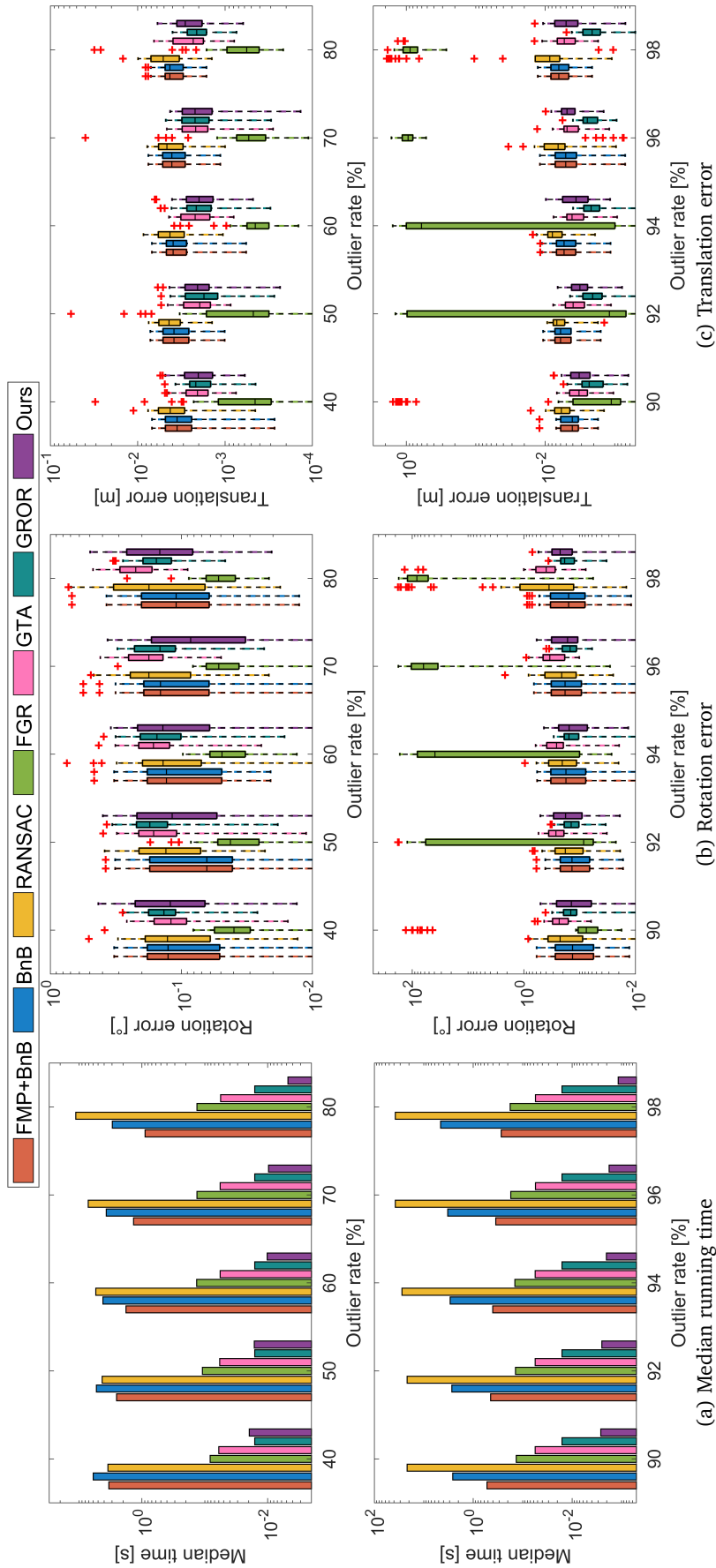


Figure 5.19: Controlled experiments on the outlier rate. (a) Median running time, (b) Rotation error, and (c) Translation error.



**Table 5.3:** Controlled experiments on the number of correspondences in the extreme case. The results include median running time (s) | success rate (%) with cases satisfying  $RE \leq 1^\circ$  and  $TE \leq 0.01m$ .

Method	Number of correspondence ( $\times 10^3$ )						
	10	20	50	100	200	500	1000
FMP+BnB	7.375 100	26.73 100	165.9 100	647.7 100	> 1800s	-	-
BnB	32.20 100	83.85 100	374.8 100	1169 100	> 1800s	-	-
RANSAC	1474 94.0	> 1800s	-	-	-	-	-
FGR	0.725 50.0	1.875 56.0	5.613 48.0	7.540 72.0	20.66 68.0	44.94 62.0	100.5 62.0
GTA	1.287 96.0	4.232 0.00	Out of memory	-	-	-	-
GROR	0.217 100	0.835 100	4.866 100	19.28 100	77.04 100	494.3 100	> 1800s
Ours	0.009 100	0.019 100	0.039 100	0.097 100	0.241 100	1.064 100	1.905 100

fixed as  $[0, 0, 1]^T$ . Therefore, the gravity direction utilized in the proposed method can be set as  $[0, 0, -1]^T$ . Subsequently, a subset of points in the target point cloud is substituted with randomly generated points in  $[-1, 1]^3$ , imitating outliers. The outlier rate, denoted by  $\eta$ , represents the proportion of these substituted points relative to the total number of points. The noise is simulated by adding zero mean Gaussian noise to both source and target point clouds, and the standard derivation is  $\sigma = 0.005$ . Following [Yan+22a; Li+23b], the inlier threshold for each method is determined by  $\sigma$ .

**Controlled experiments with different outlier rates.** This section presents two groups of controlled experiments designed to compare the outlier-robustness of the proposed three-stage method with FMP+BnB, BnB, RANSAC, FGR, GTA, and GROR. The first group of experiments has normal outlier rates, varying from 40% to 80% in increments of 10%. The second group of experiments has extremely high outlier rates, ranging from 90% to 98% in increments of 2%. The correspondence number in both experiments is fixed as  $N = 2000$ . The experiment is repeated 50 times for each setting and each method. The median running time, rotation error, and translation error are given in Fig. 5.19.

In the first group of experiments, all compared methods are robust against up to 80% outlier rate. In terms of rotation and translation accuracy, the proposed method demonstrates comparable performance to the other methods. However, Ours has the lowest running time except at 40% outlier rate. Notably, the running time of our method gradually decreases as the outlier rate increases. This phenomenon can be attributed to the reason that in the first stage, a significant number of outlier correspondences are rejected. Among the three stages, the BnB-based second stage is relatively time-consuming (worst-case exponential time complexity). Consequently, as the outlier rate increases, the input correspondences for the second stage gradually decrease and so does the total running time. This indicates the effectiveness of the proposed first stage in terms of outlier removal. FMP+BnB has a similar phenomenon as Ours and is more efficient than pure BnB since it also contains an efficient preprocessing step for outlier removal. Nonetheless, the SOTA FMP+BnB is approximately two orders of magnitude slower than Ours. The running time of RANSAC increases exponentially with the outlier rate, making it the slowest method starting from 60% outlier rate. FGR and GTA perform well under regular outlier rates. Another SOTA GROR is relatively efficient but is about

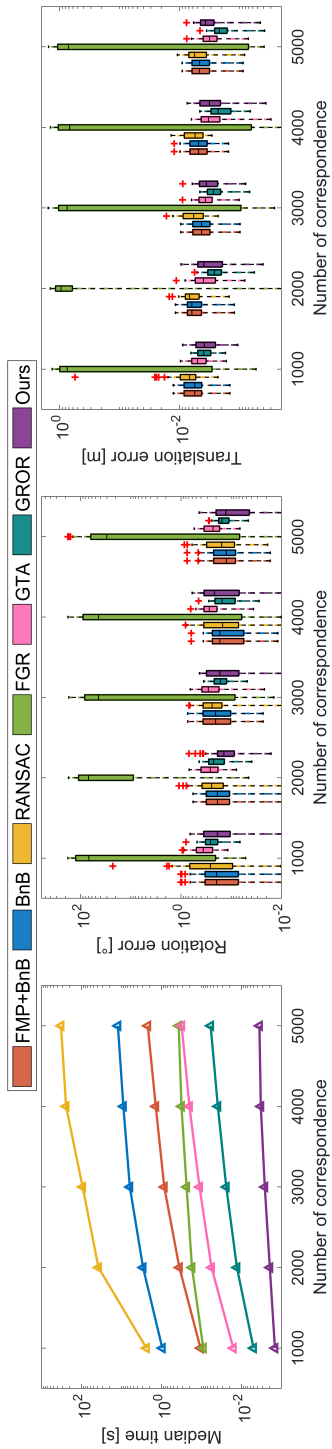


Figure 5.20: Controlled experiments on the number of correspondences. (a) Median running time, (b) Rotation error, and (c) Translation error.

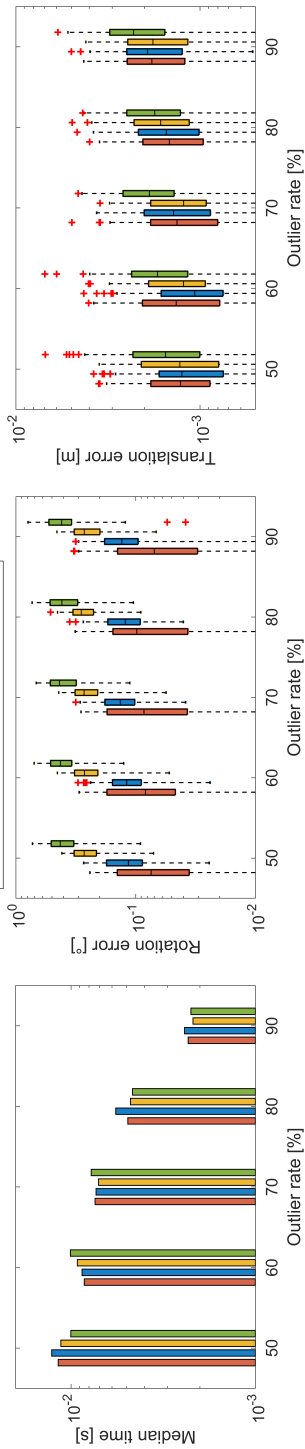


Figure 5.21: Controlled experiments on the gravity direction noise. (a) Median running time, (b) Rotation error, and (c) Translation error.

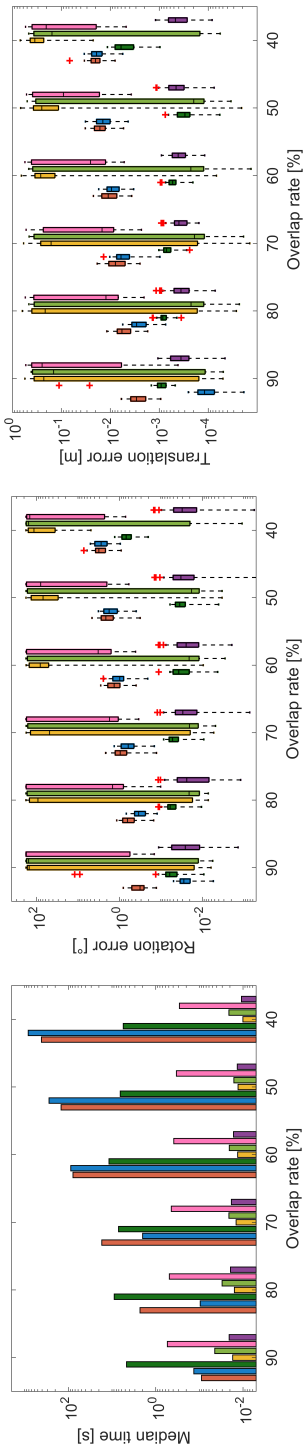
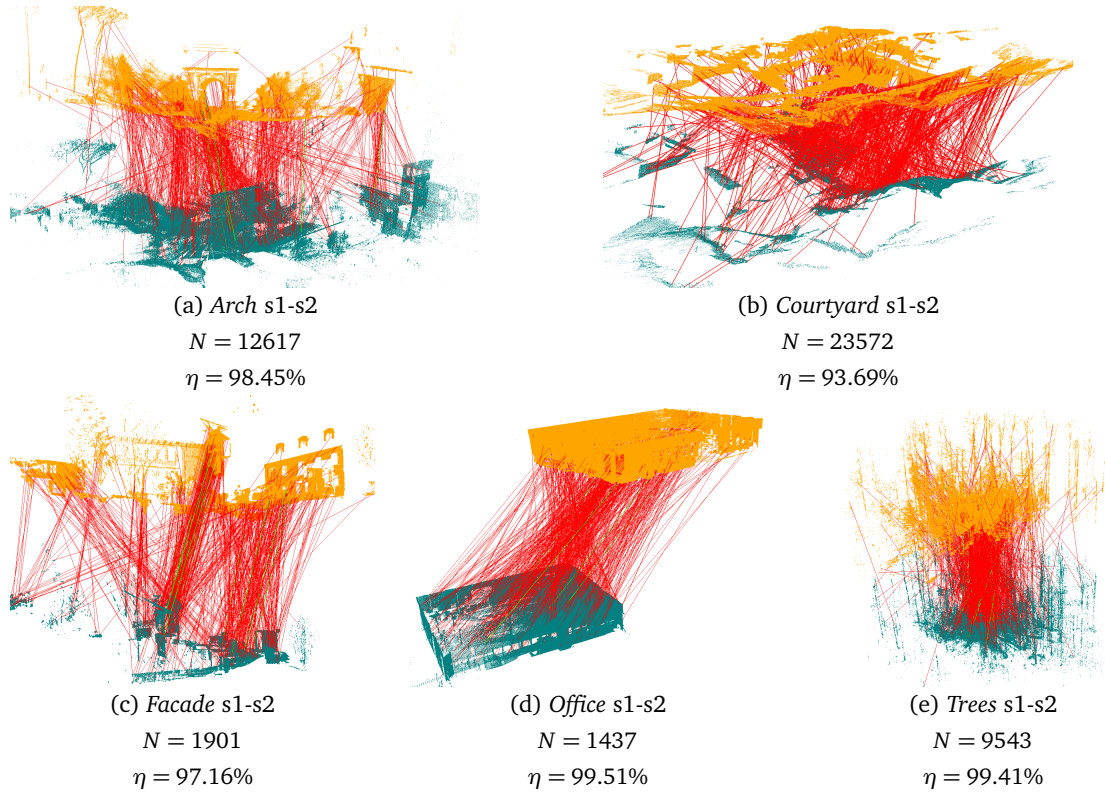


Figure 5.22: Controlled SPCR experiments on the overlap rate. (a) Median running time, (b) Rotation error, and (c) Translation error.

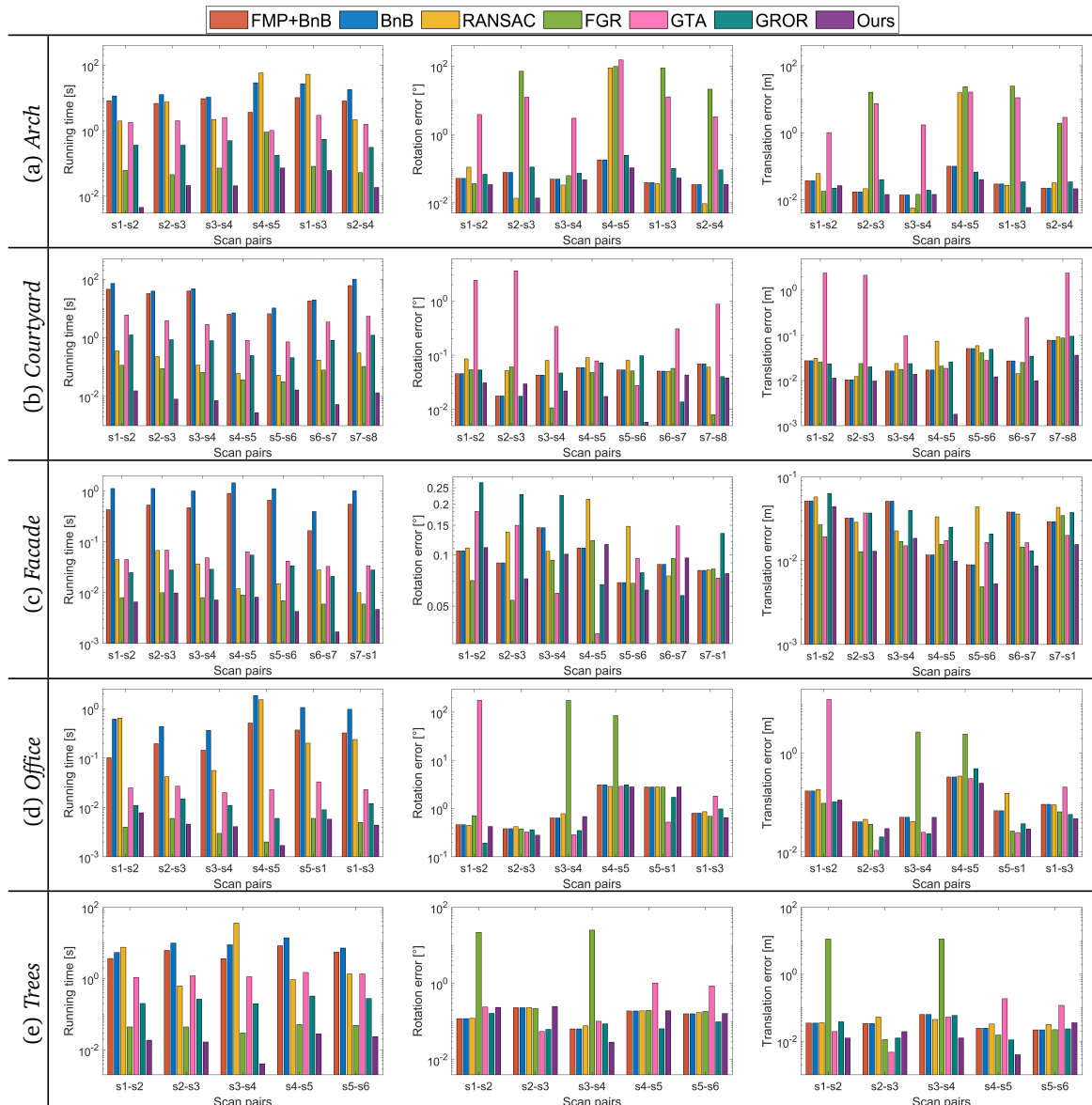


**Figure 5.23:** Examples of scan pairs and initial correspondences for each scene from ETH dataset. The number of correspondences and outlier rate are denoted by  $N$  and  $\eta$ , respectively.

3 times slower than Ours at 80% outlier rate.

In the second group of experiments, only FMP+BnB, BnB, GROR, and Ours are robust against up to 98% outlier rate while maintaining comparable registration accuracy. RANSAC only resists 94% outlier rate, and FGR starts breaking at 90% outlier rate. Due to its high number of iterations ( $10^7$ ), RANSAC is more robust than FGR but is the most time-consuming method. GTA performs relatively better but is only robust to 96% outlier rate. On the other hand, Ours outperforms all methods in terms of efficiency. For instance, Ours is approximately  $3 \times 10^4$  times faster than RANSAC, 200 times faster than FMP+BnB, and 10 times faster than GROR at 98% outlier rate. In general, Ours stands out as the fastest method and one of the most robust.

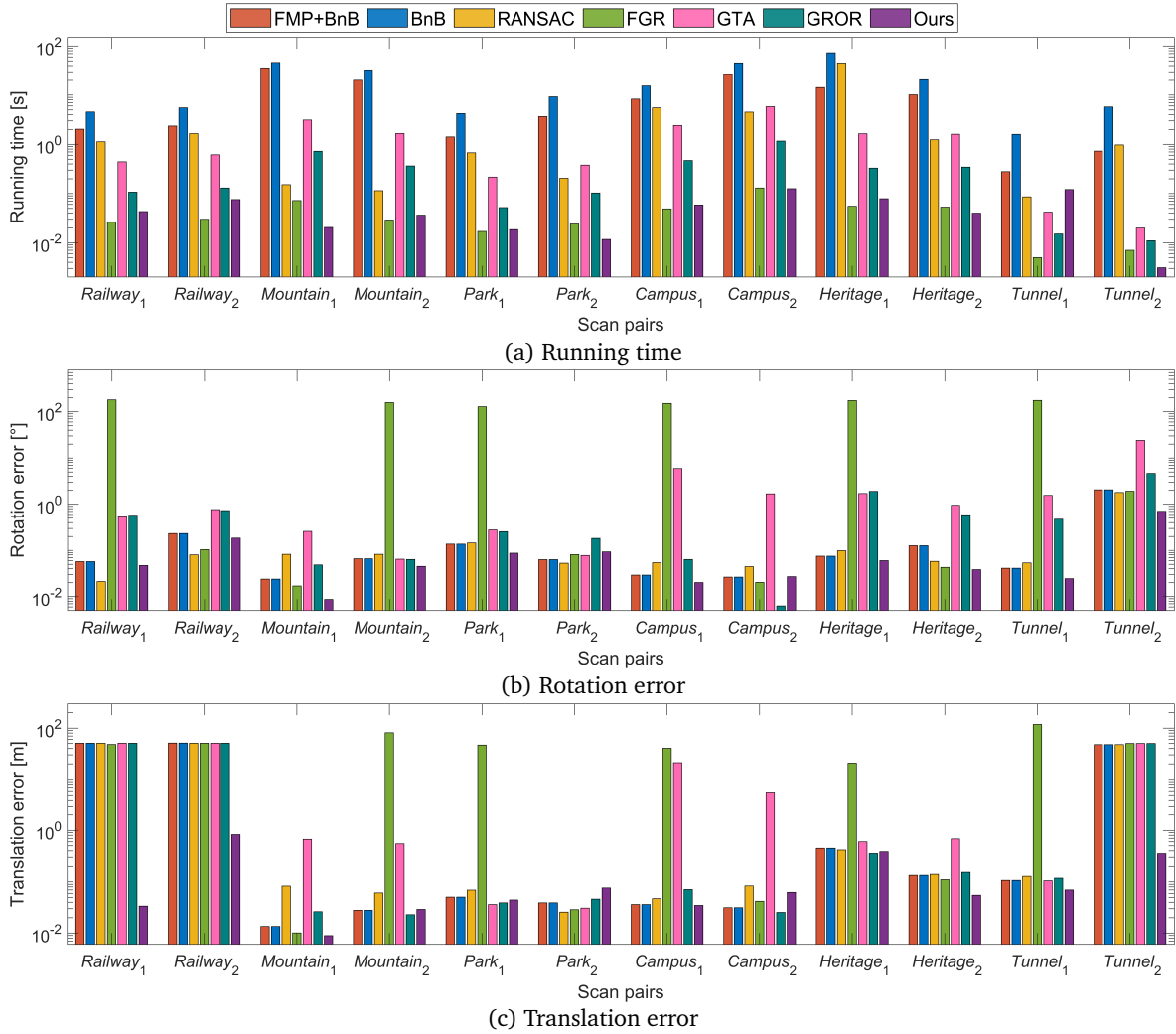
**Controlled experiments with different correspondence numbers.** This section presents two groups of controlled experiments aimed at comparing the registration efficiency. The correspondence number in the first group of experiments varies from 1000 to 5000. The outlier rate is fixed as 95%, and each experiment is repeated 50 times for each setting. The experimental results are plotted in Fig. 5.20. Compared to other methods, the efficiency of the proposed three-stage method is less affected by the correspondence number. This implies that, with an increase in the number of correspondences, the efficiency advantage of Ours will become more prominent. Similarly, the running time of FGR exhibits a slow increase as the number of correspondences grows, resulting in only a marginal time difference between GTA and FGR at  $N = 5000$ . However, FGR often produces unsatisfactory registration



**Figure 5.24:** Running times, rotation errors, and translation errors of all registration methods on the ETH dataset. (a)-(e) show the registration results for *Arch*, *Courtyard*, *Facade*, *Office*, *Trees*.

results when outlier rates are high. RANSAC is the method most sensitive to the number of correspondences and consistently exhibits the longest running time among all the methods. Particularly, as the number of correspondences increases from 1000 to 5000, the efficiency of Ours increases from being roughly 3 times faster than the second fastest method, GROR, to being 16 times faster. In the case of FMP+BnB, this improvement ranges from being 70 times faster to 613 times faster.

To further explore the scalability of Ours, the second group of experiments has extremely high correspondence numbers, varying from 10k to 1000k, where  $k$  represents  $10^3$ . This experiment also has 50 trials for each setting, and the outlier rates remain at 95%. The results, including median running times and success rates, are presented in Table 5.3. The thresholds for successful registration are  $RE \leq 1^\circ$  and  $TE \leq 0.01\text{m}$ . Notably, if the running time of a method exceeds 1800s, the results for that method are not reported. As can be seen



**Figure 5.25:** (a) Running time, (b) Rotation error, and (c) Translation error of all registration methods on the WHU-TLS dataset.

from the results of the success rate, only FMP+BnB, BnB, GROR, and Ours can successfully register all point cloud pairs even with a large scale of correspondences. However, both FMP+BnB and BnB surpass 1800s when the number of points reaches 200k. Although FGR demonstrates better efficiency, it exhibits a low success rate. As  $N$  increases to 50k, GTA is unable to even operate properly due to the memory problem. Furthermore, the running time of GROR grows faster than that of FGR as the number of correspondences increases. Therefore, GROR becomes slower than FGR starting from  $N = 100k$ , exceeding the time limit of 1800s at  $N = 1000k$ . However, thanks to the proposed transformation strategy, the results indicate that our three-stage method has the best scalability among all methods. It can successfully align 100k correspondences in 0.1s and register 1000k correspondences in 2s.

**Robustness experiments for gravity direction noise.** In practical scenarios, the gravity directions extracted from IMU measurements may exhibit bias. Consequently, we evaluate the robustness of our three-stage method against noise in gravity directions in this section.

Biased angles on gravity directions are intentionally introduced in various directions to simulate IMU noise. We then conduct controlled experiments with different biased angles, including  $0^\circ$ ,  $0.1^\circ$ ,  $0.3^\circ$ , and  $0.5^\circ$ . The experiment maintains a fixed correspondence number of  $N = 2000$ , while the outlier rate varies from 50% to 90%. We repeat 50 trials for each experimental setting. The obtained results, encompassing median running time, rotation error, and translation error, are given in Fig. 5.21. Experiment results show that varying biased angles exhibit a discernible impact on rotation error but less effect on translation error. Furthermore, the computational cost of the proposed method exhibits minimal sensitivity to gravity direction noise. In conclusion, despite gravity direction noise, the proposed three-stage method shows significant robustness against outliers and high efficiency.

**Controlled correspondence-free experiments.** In this section, we perform challenging correspondence-free experiments to compare our extended method with global GO-ICP, GO-ICPT, and local ICP, CPD, and Gmmreg, using the *Bunny* dataset from the Stanford 3-D Scanning Repository [CL96]. Due to the limitations of GO-ICP [Yan+16], the point cloud is pre-normalized to fit within the cube  $[-1, 1]^3$ . The *Bunny* dataset is initially downsampled to  $M = 234$  points, serving as the source point cloud. Random rotations and translations are then applied to generate the target point clouds. Additionally, a specific proportion of points is randomly removed from the target point cloud to simulate partial overlap between the source and target point clouds, with the overlap rate denoted by  $\rho$  ranging from 90% to 40%. Zero-mean Gaussian noise with  $\sigma = 0.001$  is added to the target point cloud. The experiments are repeated 30 times at each overlap rate, and the results are presented in Fig. 5.22. Notably, the time costs for constructing distance transform (DT) [Yan+16] in GO-ICP and GO-ICPT are not recorded but averaged about 12s.

It can be observed from the rotation and translation errors that local ICP tends to converge to local minima even at  $\rho = 90\%$ . Due to the adoption of GMM, the local CPD and Gmmreg are more robust to partial overlap than ICP, as evidenced by their better error distributions than those of ICP. Notably, global GO-ICP and GO-ICPT are much more robust than these local methods. Among them, GO-ICPT (50%) demonstrates the best performance, indicating that GO-ICP is strongly sensitive to the pre-set trimming percentage. However, as the overlap rate decreases, the time costs of GO-ICP and GO-ICPT increase significantly compared to local methods. In contrast, the proposed global method attains commendable robustness and accuracy even at  $\rho = 40\%$ . On the other hand, it is faster than the existing global method (i.e., GO-ICP) and even has comparable efficiency to local methods. This shows the feasibility of our three-stage method in addressing correspondence-free registration problems.

### Real-World Data Experiments

To evaluate the registration performance on the real-world data, we use three challenging large-scale datasets, including ETH dataset [TWS14] and WHU-TLS dataset [Don+20]. Following the preparation strategy in [Li22; Li+23b], the first step is to downsample the original point cloud using the voxel grid algorithm [RC11]. The second step is to extract ISS [Zho09]

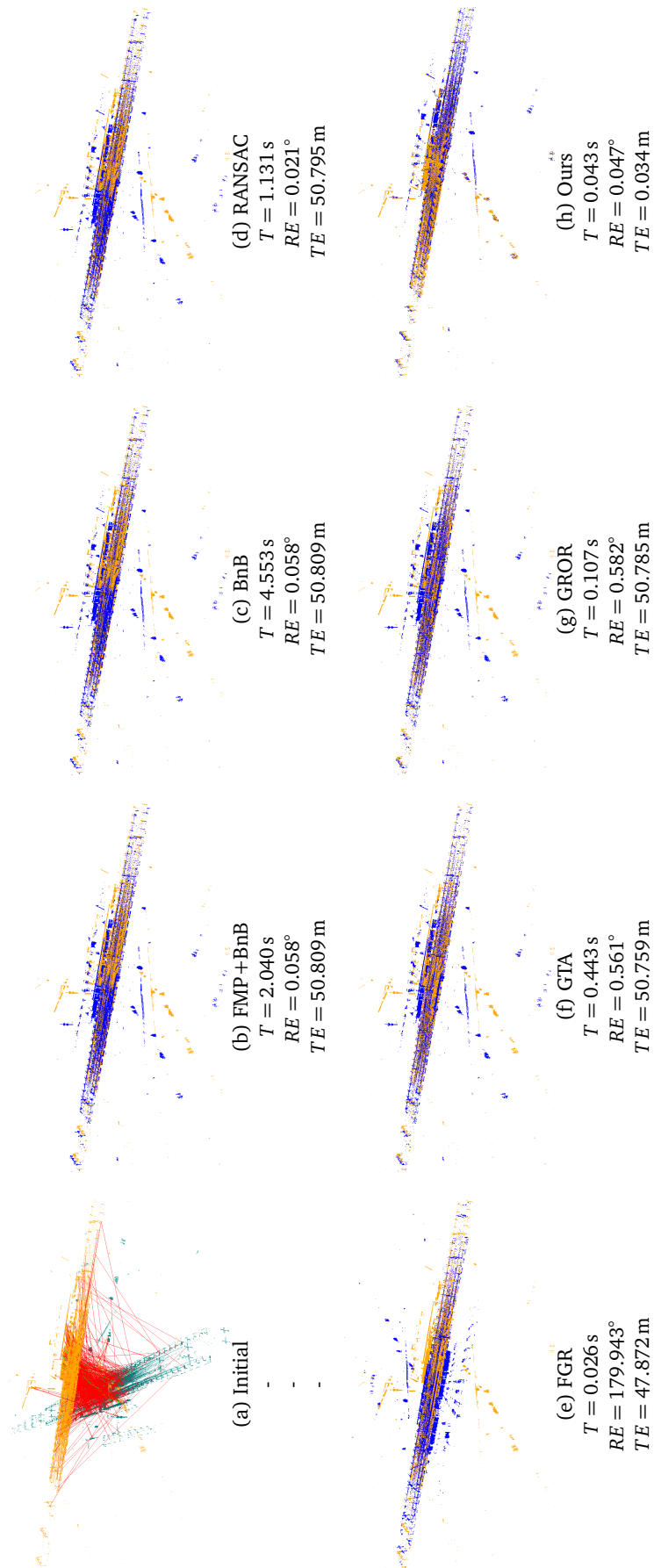


Figure 5.26: Qualitative and quantitative registration results for scan pair *Railway<sub>1</sub>* on the WHU-TLS dataset.



**Table 5.4:** Detailed information about the ETH dataset

Scans	Number of points ( $\times 10^6$ )	Number of keypoints	Number of correspondences	Outlier rate	
<i>Arch</i>	s1-s2	23.56-30.90	19007-12254	12617	98.45%
	s2-s3	30.90-25.25	12254-19192	13217	99.33%
	s3-s4	25.25-29.45	19192-13286	15187	98.54%
	s4-s5	29.45-27.96	13286-11838	9312	99.77%
	s1-s3	27.96-25.25	19007-19192	16608	99.71%
	s2-s4	30.90-29.45	12254-13286	11699	98.77%
<i>Courtyard</i>	s1-s2	13.32-11.45	12450-26189	23572	93.69%
	s2-s3	11.45-12.71	26189-9634	19018	92.33%
	s3-s4	12.71-12.15	9634-12125	15325	86.55%
	s4-s5	12.15-16.75	12125-4081	8069	90.62%
	s5-s6	16.75-11.99	4081-11208	7534	93.28%
	s6-s7	11.99-11.17	11208-13338	17369	92.22%
	s7-s8	11.17-11.45	13338-20386	22649	95.37%
<i>Facade</i>	s1-s2	25.07-15.25	1586-2810	1901	97.16%
	s2-s3	15.25-15.79	2810-2215	2368	96.92%
	s3-s4	15.79-23.48	2215-1950	1959	96.43%
	s4-s5	23.48-16.74	1950-1845	2204	93.06%
	s5-s6	16.74-21.46	1845-1576	1809	94.25%
	s6-s7	21.46-20.79	1576-1555	1615	96.78%
	s7-s1	20.79-25.07	1555-1586	1601	94.25%
<i>Office</i>	s1-s2	10.72-10.71	1355-1407	1437	99.51%
	s2-s3	10.71-10.73	1407-1348	1474	97.15%
	s3-s4	10.73-10.69	1348-1277	1279	97.65%
	s4-s5	10.69-10.75	1277-1486	1355	98.97%
	s5-s1	10.75-10.72	1486-1355	1616	97.34%
	s1-s3	10.72-10.73	1355-1348	1382	98.77%
	s3-s5	10.73-10.75	1348-1486	1578	98.23%
<i>Trees</i>	s1-s2	19.63-19.60	10883-10898	9543	99.41%
	s2-s3	19.60-19.77	10898-10873	9897	97.61%
	s3-s4	19.77-20.39	10873-12542	9811	99.76%
	s4-s5	20.39-20.48	12542-12522	11253	97.87%
	s5-s6	20.48-21.61	12522-11894	10711	98.43%

keypoints and calculate FPFH [RBB09] descriptors for each keypoint. The third step is establishing the set of putative correspondences  $\mathcal{C}$  through the K-nearest neighbor search [Low04].

**ETH dataset experiments.** The ETH dataset is a large-scale terrestrial LiDAR dataset, which contains 5 scenes: *arch*, *courtyard*, *facade*, *office*, and *trees*. In each challenging scene, sev-



**Table 5.5:** Detailed information about the WHU-TLS dataset

Scans	Number of points ( $\times 10^6$ )	Number of keypoints	Number of correspondences	Outlier rate
<i>Railway</i> <sub>1</sub>	6.74-6.34	6399-5180	6013	99.60%
<i>Railway</i> <sub>2</sub>	6.76-6.74	6500-6399	7073	99.80%
<i>Mountain</i> <sub>1</sub>	3.75-3.68	13287-11584	16243	89.42%
<i>Mountain</i> <sub>2</sub>	3.68-3.46	11584-8554	11805	92.75%
<i>Park</i> <sub>1</sub>	5.66-5.07	4849-5880	4248	98.68%
<i>Park</i> <sub>2</sub>	5.07-5.76	5880-7055	5564	97.27%
<i>Campus</i> <sub>1</sub>	12.44-11.68	18087-21353	14885	99.10%
<i>Campus</i> <sub>2</sub>	11.68-8.75	21353-40446	23591	98.68%
<i>Heritage</i> <sub>1</sub>	28.50-26.19	13827-14942	12019	99.89%
<i>Heritage</i> <sub>2</sub>	31.48-28.50	13409-13827	11645	98.38%
<i>Tunnel</i> <sub>1</sub>	22.36-22.42	1954-2098	1871	98.18%
<i>Tunnel</i> <sub>2</sub>	22.42-22.40	1799-1629	1319	99.70%

**Table 5.6:** Quantitative registration results on real-world datasets. The results include average running time (s) | average rotation error ( $^\circ$ ) | average translation error (m). Bolded fonts indicate the best results.

Method	Dataset					
	<i>ETH-Arch</i>	<i>ETH-Courtyard</i>	<i>ETH-Facade</i>	<i>ETH-Office</i>	<i>ETH-Trees</i>	<i>WHU-TLS</i>
FMP+BnB	7.722 0.072 0.037	30.02 0.048 0.033	0.532 0.099 0.032	0.276 1.368 0.126	5.485 0.153 0.036	10.45 0.245 12.51
BnB	18.12 0.072 0.037	42.33 0.048 0.033	1.035 0.099 0.032	0.885 1.368 0.126	9.099 0.153 0.036	22.02 0.245 12.51
RANSAC	20.71 14.87 3.139	0.185 0.071 0.044	0.031 0.124 0.038	0.451 1.369 0.144	9.327 0.159 0.040	5.121 0.214 12.52
FGR	0.202 46.31 11.07	0.075 0.041 0.035	0.008  <b>0.084</b>  0.018	<b>0.004</b>  44.16 0.899	0.044 9.533 4.532	<b>0.041</b>  80.66 37.96
GTA	1.966 31.12 6.747	3.320 1.093 1.056	0.048 0.106 0.020	0.025 30.22 2.155	1.253 0.458 0.077	1.499 3.161 15.09
GROR	0.374 0.116 0.036	0.777 0.049 0.039	0.031 0.151 0.034	0.011  <b>1.122</b>  0.122	0.253  <b>0.096</b>  0.029	0.318 0.795 12.72
Ours	<b>0.033</b>   <b>0.048</b>   <b>0.020</b>	<b>0.010</b>   <b>0.027</b>   <b>0.014</b>	<b>0.006</b>  0.091  <b>0.016</b>	0.005 1.278  <b>0.086</b>	<b>0.018</b>  0.174  <b>0.017</b>	0.053  <b>0.112</b>   <b>0.165</b>

eral scans are captured from different positions, thus suffering from low overlap, self-similar structures, and occlusion. As visualized in Fig. 5.23, the outlier rates for various scan pairs in the ETH dataset are notably high, ranging approximately from 86% to 99%. We employ adjacent scans as the input of pair-wise registration in each scene. Following [Yan+22a], the downsample resolution is set to 0.1 m, as well as the inlier threshold. The detailed information can be found in Table 5.4, including the number of points, number of keypoints, number of correspondences, and outlier rate. Notably, the gravity direction employed for the proposed method is set to  $[0, 0, -1]^T$ . Fig. 5.24 shows the alignment results for a total of 32 scan pairs, and Table 5.6 presents the average running times and average errors for each scene.

As can be seen from the results, only FMP+BnB, BnB, GROR, and Ours successfully register all scan pairs with relatively low errors. Notably, with the exception of *Office*, our method demonstrates the lowest average time cost in most scenes. Besides, Ours achieves the lowest average rotation and translation errors in *Arch* and *courtyard*, as well as the lowest average translation errors in *facade*, *office*, and *trees*. Although FMP+BnB and BnB perform admirably across all scenes, they are computationally expensive compared to other methods. RANSAC,

**Table 5.7:** Detailed information about the real-world datasets in SPCR experiments.

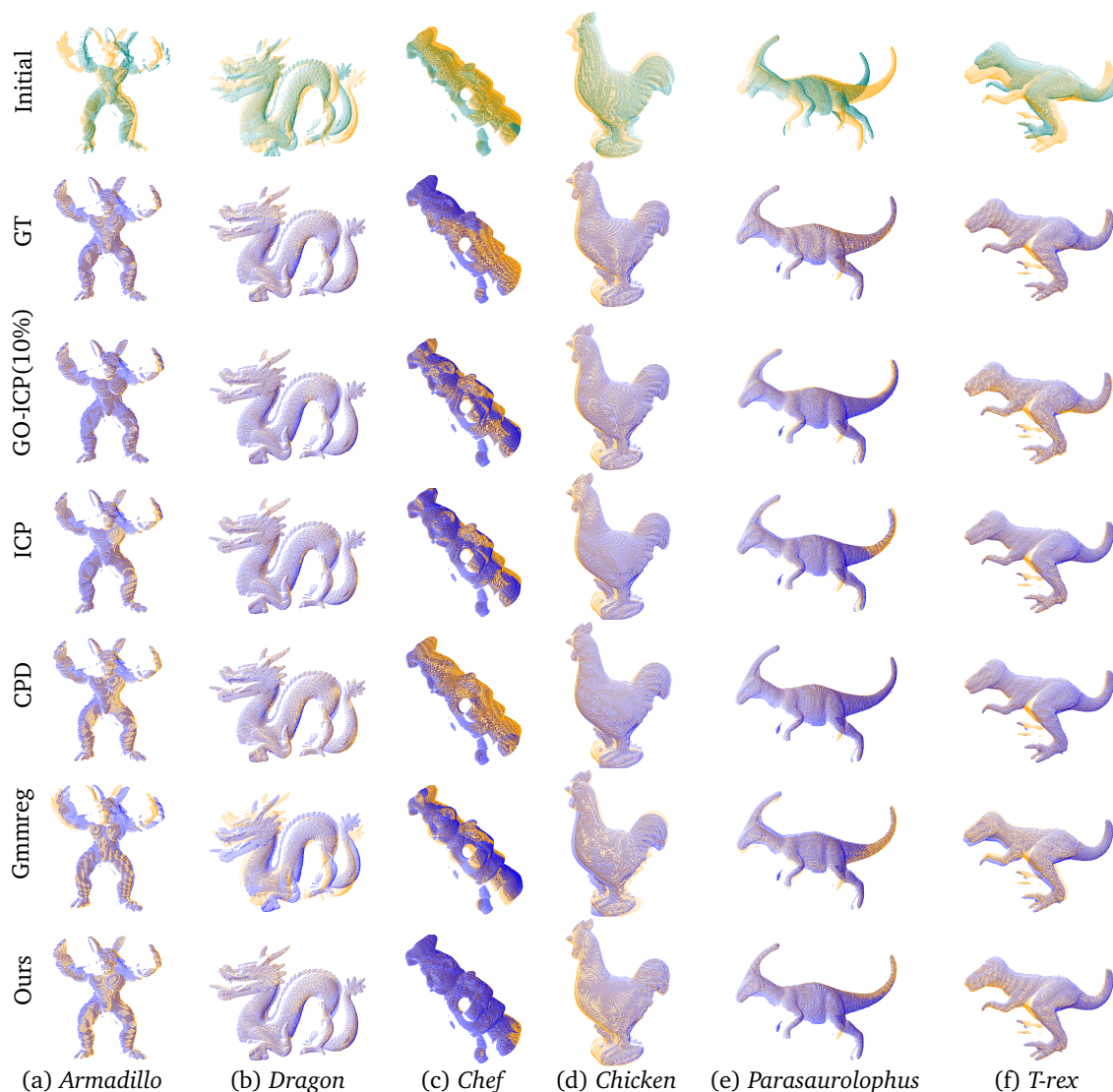
Dataset	Downsample resolution	Number of points ( $\times 10^3$ )	Number of points (after downsample)
Stanford-Armadillo	0.006	26.94-25.57	865-801
Stanford-Dragon		41.84-34.84	1081-953
U3M-Chef		66.87-65.96	1358-1286
U3M-Chicken		29.52-30.17	590-630
U3M-Parasaurolophus		46.00-42.94	944-849
U3M-Trex		38.78-40.21	771-863

**Table 5.8:** Quantitative correspondence-free registration results on real-world datasets, including running time (s) | rotation error ( $^\circ$ ) | translation error (m). Bolded fonts indicate the best results.

Method	Dataset					
	Stanford-Armadillo	Stanford-Dragon	UWA-Chef	UWA-Chicken	UWA-Parasaurolophus	UWA-Trex
GO-ICP	12.89 4.745 0.003	12.85 0.406  <b>0.001</b>	13.18 6.294 0.070	12.97 1.931 0.022	12.96 4.278 0.047	13.15 1.273 0.014
GO-ICPT(10%)	12.98 0.751  <b>0.001</b>	12.98 0.276  <b>0.001</b>	13.04 4.619 0.052	12.86 0.987 0.010	13.19 0.613 0.005	12.88 0.692  <b>0.003</b>
GO-ICPT(50%)	13.02  <b>0.193</b>   <b>0.001</b>	12.94 0.205  <b>0.001</b>	12.98 14.20 0.158	13.01 19.42 0.214	12.99 1.301 0.008	12.95 9.853 0.108
ICP	<b>0.045</b>  4.393 0.003	<b>0.027</b>  0.415  <b>0.001</b>	<b>0.042</b>  6.058 0.068	<b>0.020</b>  1.691 0.019	<b>0.020</b>  4.288 0.047	<b>0.014</b>  1.220 0.014
CPD	0.547 1.126  <b>0.001</b>	0.547 0.456  <b>0.001</b>	4.651 1.968 0.021	0.234 3.765 0.043	0.611 0.300  <b>0.002</b>	0.520  <b>0.518</b>  0.006
Gmmreg	2.730 3.818 0.004	1.941 12.07 0.025	3.547 14.73 0.164	1.592 7.783 0.087	6.748 2.192 0.020	2.850 4.625 0.049
Ours	0.187 0.325  <b>0.001</b>	0.394  <b>0.175</b>   <b>0.001</b>	0.302  <b>1.224</b>   <b>0.018</b>	0.083  <b>0.433</b>   <b>0.001</b>	0.133  <b>0.216</b>  0.005	0.116 0.585 0.016

owing to its non-deterministic nature, occasionally produces inaccurate results, as seen in the case of  $s_4$ - $s_5$  in *Arch*. Additionally, RANSAC becomes considerably time-consuming when dealing with outlier rates surpassing 99%, as evident in the registration of scan pairs  $s_4$ - $s_5$  ( $\eta = 99.77\%$ ) and  $s_1$ - $s_3$  ( $\eta = 99.71\%$ ) in *Arch*. While FGR is highly efficient, it is susceptible to failure, particularly in scenes with high outlier rates. A total of 8 scan pairs exhibit significant registration errors in the results obtained by FGR. Similar to RANSAC, GTA is efficient but also occasionally generates erroneous results due to its non-deterministic nature. This is particularly evident in *Arch*, where GTA exhibits high registration errors, including a  $31.12^\circ$  average rotation error and a 6.747m average translation error. GROR is the second-best approach, which has high accuracy and high efficiency. Overall, benefiting from the proposed transformation decoupling strategy, our method shows superior registration efficiency and accuracy compared to SOTA methods.

**WHU-TLS dataset experiments.** The WHU-TLS dataset is another large-scale terrestrial LiDAR dataset. To ensure the generality of the registration algorithm, we randomly select two scan pairs each from the *Railway*, *Mountain*, *Park*, *Campus*, *Heritage* and *Tunnel* scenes for the registration experiment. Following [Yan+22a], the downsample resolution is set to 0.2m, as well as the inlier threshold. Detailed information about the selected scan pairs is given in Table 5.5. The outlier rates for these chosen pairs range from approximately 89% to 99%, with correspondence numbers spanning from around 1k to 20k. Besides, the gravity direction remains fixed at  $[0, 0, -1]^T$ . The registration results for a total of 12 scan pairs, including running time, rotation error, and translation error, are shown in Fig. 5.25. In



**Figure 5.27:** Qualitative correspondence-free registration results on the Stanford and UWA datasets. (a) Stanford-Armadillo, (b) Stanford-Dragon, (c) U3M-Chef, (d) U3M-Chicken, (e) U3M-Parasaurolophus, (f) U3M-T-rex. The source, target, and aligned point clouds are green, orange, and blue.

addition, the quantitative average results are presented in Table 5.6.

Observing the average rotation and translation errors, it is evident that only our method successfully registers all scan pairs in the WHU-TLS dataset. Despite their robust performance in the ETH dataset, both FMP+BnB and BnB, as well as GROR, exhibit significant translation errors when registering scan pairs  $Railway_1$ ,  $Railway_2$ , and  $Tunnel_2$ . This phenomenon can be attributed to the abundance of self-similar structures frequently appearing in these railway and tunnel scenes. As shown in Fig. 5.26, we provide qualitative and quantitative registration results for scan pair  $Railway_1$  to further illustrate this phenomenon. All methods, except Ours, converge to similar registration solutions (local minima), as their translation errors are nearly identical. This demonstrates the robustness of our approach to the abundant outliers generated by practical self-similar structures. Similar to the findings in the ETH dataset experiment, Ours, along with FMP+BnB, BnB, and GROR, achieve good registration results for scan pairs other than those mentioned above. The performance of FGR and GTA is unsta-

ble, as they occasionally yield unsatisfactory results despite their high efficiency. Besides, RANSAC has high computational costs second only to FMP+BnB and BnB in most scenes. In contrast, the proposed three-stage method not only achieves excellent accuracy (with the lowest average rotation and translation errors) but also achieves the second-highest computational efficiency, surpassed only by FGR.

**Correspondence-free experiments.** In order to further assess the performance of the extended correspondence-free registration method, we employ benchmark datasets, i.e., the Stanford dataset [CL96] and UWA dataset [MBO06]. A comprehensive comparison is conducted against baseline methods, including GO-ICP, GO-ICPT(10%), GO-ICPT(50%), ICP, CPD, and Gmmreg. In the experiment, we use the *Armadillo* and *Dragon* models from the Stanford dataset, as well as the *Chef*, *Chicken*, *Parasaurolophus*, and *T-rex* models from the UWA dataset. These datasets exhibit characteristics of partial overlap and self-occlusion. The methods are tested on one pair of partially overlapped point clouds constructed from each model, with a downsample resolution set to 0.006. The details on these scan pairs are presented in Table 5.7. The registration results are listed in Table 5.8, and qualitative outcomes are visualized in Fig. 5.27.

From the results, the proposed deterministic global method obtains the lowest registration error across most models. Consistent with the findings in synthetic data experiments, GO-ICP exhibits strong sensitivity to the pre-set trimming percentage since GO-ICPT (10%) delivers the most favorable results among GO-ICP and GO-ICPT. Local methods, including ICP, CPD, and Gmmreg, display instability and occasionally yield unsatisfactory registration results (as seen in Fig. 5.27). Additionally, from an efficiency standpoint, our method is significantly faster than global methods while maintaining comparable efficiency to local methods. These findings also showcase the feasibility of the proposed method in addressing correspondence-free registration challenges.

### 5.4.3 Registration of LiDARs in ITS Infrastructure

To validate and compare the performance of the proposed two-stage and three-stage methods in autonomous driving scenarios, we employ the A9 dataset [Cre+22], which is gathered by multiple roadside sensors mounted on the ITS Infrastructure, as shown in Fig. 5.28. The purpose of this experiment is to align point clouds collected from two distinct LiDARs. In particular, the captured point clouds contain diverse moving objects (as shown in Fig. 5.29), including cars, buses, and pedestrians, resulting in the challenging registration task. We select 5 scan pairs with a minimum temporal separation of 10s as input for the experiment. Following the segmentation of ground points, we leverage non-ground points to establish putative correspondences utilizing the same generation strategy in the last section. The downsample resolution is set to 0.1 m. Nevertheless, the outlier rates for both scan pairs exceed 99%. Table 5.9 provides detailed information about the selected scan pairs.

The proposed two methods are compared with 5 baselines: FMP+BnB [Cai+19], RANSAC [FB81], FGR [ZPK16], GTA [ART10], and GROR [Yan+22a]. In addition, the two-stage and three-



**Figure 5.28:** One of the measurement points of the Providentia++ project on the A9 highway (<https://innovation-mobility.com/en/project-providentia/a9-dataset/>). The A9 dataset [Cre+22] is obtained from roadside sensors, including multiple LiDARs, cameras, and radars.

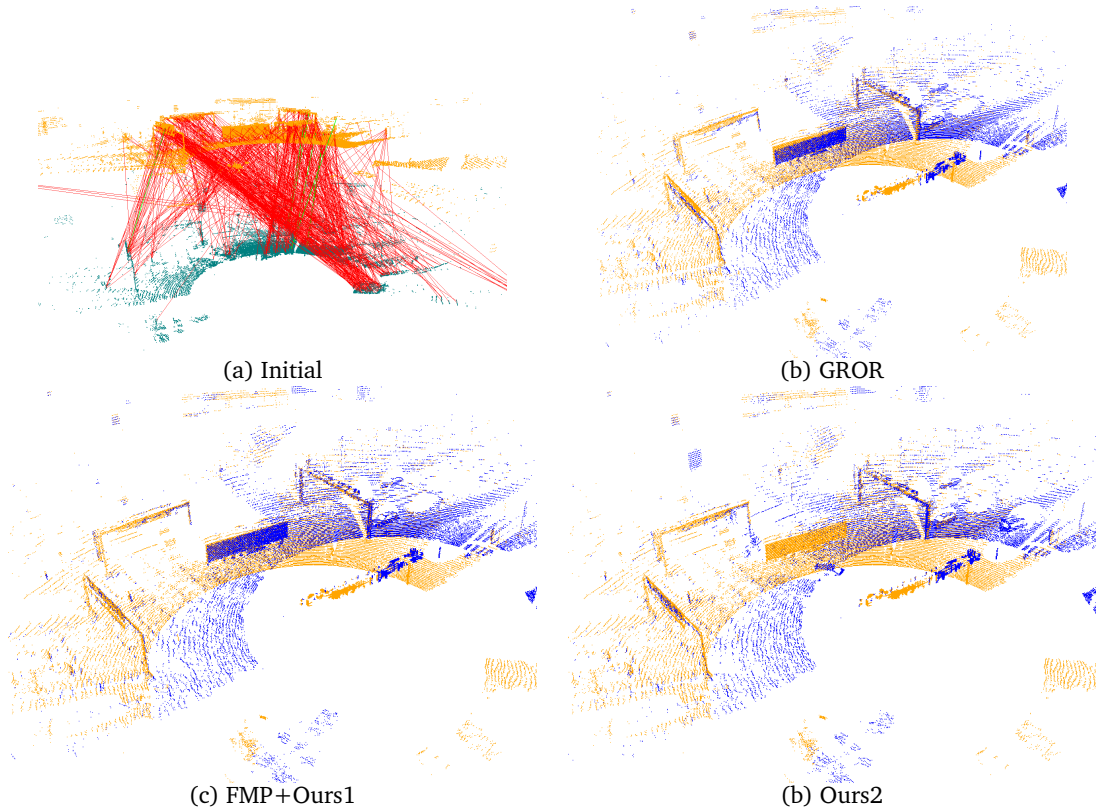
**Table 5.9:** Detailed information about the A9 dataset

Scans	Number of points ( $\times 10^6$ )	Number of keypoints	Number of correspondences	Outlier rate
P1	0.114-0.110	1099-1392	2883	99.06%
P2	0.114-0.110	801-1443	2327	99.18%
P3	0.113-0.110	983-1401	2657	99.26%
P4	0.114-0.109	840-1183	2359	99.49%
P5	0.113-0.110	964-1124	2410	99.21%

stage methods are referred to as **Ours1** and **Ours2**, respectively. Since both LiDARs are parallel to the ground, the gravity direction employed remains  $[0, 0, -1]^T$ . The inlier thresholds are set to 0.1m and  $0.5^\circ$  for distance-based and angle-based methods. In the process of operating Ours1 and BnB, FMP is used to accelerate the convergence of BnB-based methods. Additionally, all experiments are implemented in a PC with an AMD 5600x CPU and 32GB RAM.

The registration results of all methods are shown in Table 5.10. Qualitative comparison on scan pair P3 is provided in Fig. 5.29. The experimental results demonstrate that the proposed two methods successfully register all scan pairs with significant accuracy. FMP+BnB also delivers accurate results for all scan pairs but is relatively expensive in computation. RANSAC is the most time-consuming method in this dataset due to the extremely high outlier rates. FGR and GTA are both very efficient. However, the performance of RANSAC, FGR, and GTA is unstable, resulting in high average registration errors. GROR is capable of aligning all scan pairs efficiently but has a relatively high average rotation error. In contrast, the proposed method FMP+Ours1 is more efficient than FMP+BnB and achieves the best translation accuracy. Additionally, the proposed method Ours2 has comparable efficiency to





**Figure 5.29:** Qualitative results for scan pair P3 on the A9 dataset. (a) Initial, (b) GROR, (c) FMP+Ours1, (d) Ours2.

GROR and is an order of magnitude faster than FMP+BnB. This demonstrates the effectiveness of the two transformation decoupling strategies in our proposed methods. Particularly, by further exploring the geometric properties, the three-stage method proposed based on the screw theory exhibits better performance than the two-stage method.

## 5.5 Summary

This section introduces two novel transformation decoupling strategies for robust point cloud registration with gravity prior. In the first approach, the known gravity directions are lever-

**Table 5.10:** Quantitative registration results on the A9 dataset. The results include running time (s) | rotation error ( $^{\circ}$ ) | translation error (m). Bolded fonts indicate the best results.

Method	Dataset						
	P1	P2	P3	P4	P5	Mean	
FMP+BnB	0.326 0.428  <b>0.265</b>	0.307 0.548 0.328	0.608 0.416 0.161	0.836 0.806 0.415	0.838 0.919 0.647	0.583 0.624 0.363	
RANSAC	20.71 0.606 0.497	10.84  <b>0.387</b>   <b>0.252</b>	10.53 0.461 0.173	13.25 0.929  <b>0.403</b>	12.40 34.41 10.44	13.55 7.359 2.354	
FGR	0.011 10.83 6.531	<b>0.006</b>  30.47 9.165	<b>0.006</b>  2.911 7.549	<b>0.009</b>  37.33 9.452	<b>0.006</b>  7.220 7.363	<b>0.008</b>  17.75 8.012	
GTA	0.101 33.93 7.734	0.065 102.4 12.16	0.083 179.8 16.01	0.071 6.089 7.723	0.070 19.16 8.735	0.078 68.28 10.47	
GROR	0.048 2.202 0.582	0.036 0.430 0.323	0.039 0.421  <b>0.090</b>	0.038 3.076 1.363	0.038 2.132 0.268	0.040 1.652 0.525	
FMP+Ours1	0.265  <b>0.325</b>  0.382	0.204 0.412 0.333	0.483 0.405 0.179	0.613 0.928 0.521	0.673 0.984  <b>0.225</b>	0.448 0.611  <b>0.328</b>	
Ours2	<b>0.005</b>  0.388 0.398	0.058 0.388 0.365	0.140  <b>0.386</b>  0.254	0.050  <b>0.412</b>  0.488	0.023  <b>0.397</b>  0.299	0.055  <b>0.394</b>  0.361	

aged to decouple the joint 4-DOF pose into the sequential 3-DOF translation and 1-DOF rotation. Then, a deterministic two-stage solver, including BnB and global voting, is developed for the two sub-problems. By leveraging screw theory to reformulate the registration problem, the second approach successfully decouples the 4-DOF registration problem into three sub-problems with 1-DOF, 2-DOF, and 1-DOF, respectively. Then, we propose an efficient and deterministic three-stage method to tackle these sub-problems, including interval stabbing, BnB, and global voting. These transformation decoupling strategies significantly enhance registration efficiency. In particular, the three-stage method further reduces the dimensionality of the parameter space, thus obtaining a better efficiency than the two-stage method. Overall, extensive experiments on both synthetic and real-world data shed light on the favorable performance of our methods in terms of efficiency and robustness.





# 6

## Conclusion

In this chapter, important conclusions are drawn based on the study carried out in this thesis, and the new possible directions for future work are outlined to address the limitations of the proposed methods.

### 6.1 Summary

Robust geometric perception of the surroundings is crucial for intelligent transportation systems to achieve complete autonomy within dynamic and complex traffic environments. Pose estimation is one of the most widely used tasks in geometric perception. This thesis explores highly robust solutions for typical pose estimation problems in ITSs, including sensor extrinsic calibration and point cloud registration. From a mathematical standpoint, the pose estimation problem, as discussed in previous studies [LH07; Yan+16], is inherently non-convex. Consequently, tackling the pose estimation problem necessitates addressing the non-convex optimization problem. Specifically, since the existence of noise and outliers is unavoidable in real-world applications, it is essential to suppress their negative effects. To address this issue, robust objective functions are formulated, typically non-convex and possibly non-smooth, enabling the distinction between inliers and outliers. Our research aims to seek the optimal solution for non-convex optimization problems, which often exhibit multiple local optima. However, relying solely on locally optimal or non-deterministic algorithms poses a significant risk of encountering failures, as they cannot guarantee the identification of the global optimum. In order to meet the safety requirements of ITS, this thesis applies the BnB framework to obtain global and robust solutions for various pose estimation problems. The presented deterministic global approaches enable us to effectively handle geometric optimization problems involving a substantial number of outliers.

## 6.2 Primary Contributions

This thesis studies the robust geometric perception problem in ITSs, specifically the pose estimation problem. Concretely, the primary contributions of this thesis are as follows:

**Chapter 3.** A globally optimal and robust extrinsic calibration method for traffic radar is proposed in this chapter. This targetless method can calibrate extrinsic parameters during ITS operation without any artificial objects. In addition, a novel BnB-based globally optimal registration method is introduced in the optimization back-end. This method effectively aligns the measurements of radar and GPS in a simultaneous pose and correspondence registration manner. Practical experiments demonstrate the robustness of the proposed method against noise and outliers in radar measurements. Furthermore, it successfully avoids failures caused by falling into local optima, particularly when dealing with large relative angles.

**Chapter 4.** An efficient and robust point cloud registration method is proposed in this chapter. Specifically, a novel pose decoupling strategy based on residual projections is introduced, effectively decomposing the raw problem into three search sub-problems. Compared with existing methods, the proposed approach searches for the optimal solution in the low-dimensional solution domain, thereby improving search efficiency. Subsequently, a novel upper bound based on interval stabbing is derived, and accordingly, a BnB-based search algorithm is proposed to solve these sub-problems. Moreover, the proposed upper bound is modified by interval merging to solve the registration problem without correspondence. Extensive experiments demonstrate that the proposed method outperforms state-of-the-art methods in terms of efficiency while ensuring robustness.

**Chapter 5.** An efficient and robust point cloud registration method with gravity prior is proposed in Section 5.2. Firstly, a novel transformation decoupling scheme is presented, which can reduce the solution domain to 3-DOF+1-DOF with the aid of known gravity directions. Subsequently, this section presents a BnB-based optimization algorithm for the 3-DOF translation search and an efficient global voting algorithm for the 1-DOF rotation estimation. Moreover, the proposed method can be applied to solve the correspondence-free registration problem using the all-to-all correspondence assumption. By further exploring the geometric constraints for point cloud registration from the perspective of screw theory, another efficient and robust method for registration with gravity prior is presented in Section 5.3. Specifically, a novel transformation decoupling strategy is proposed to decompose the 4-DOF registration problem into three sub-problems with 1-DOF, 2-DOF, and 1-DOF, respectively. Accordingly, a novel three-stage solver for the decoupled sub-problems is proposed, which contains interval stabbing, BnB, and global voting algorithms. The proposed approach is also extended to solve the challenging correspondence-free registration problem without the all-to-all correspondence assumption. Extensive experiments demonstrate the superiority of both algorithms in terms of efficiency and robustness.

## 6.3 Limitations

This thesis proposes different approaches to address robust pose estimation problems in ITS to meet the safety demand. In particular, all methods are dependent on the BnB optimization framework to obtain the global optimum of the estimation problem. For the low-dimensional issues (e.g., extrinsic calibration in Chapter 3), BnB is directly used to search for the globally optimal solutions. However, the computational complexity of BnB is exponential to the dimensionality of the solution domain in the worst-case. Hence, we develop different transformation decoupling strategies for BnB to address the high-dimensional problems (e.g., point cloud registration in Chapter 4 and Chapter 5). Despite the promising results of the proposed methods for pose estimation, there are still notable limitations:

- The proposed BnB-based methods are relatively time-consuming. While the BnB algorithm is more efficient than exhaustive search, and we develop transformation decoupling strategies to accelerate BnB, it is still far from real-time. In practice, when the outlier rate or data size of the input is large, the BnB algorithm commonly would be computationally intractable.
- The optimal solutions of the decoupled sub-problems are not necessarily globally optimal for the raw problem. The transformation decoupling is a geometric relaxation so that the decoupled sub-problems only keep partial constraints compared to the original problem.
- The proposed BnB-based methods are only suitable for low-dimensional problems. The BnB algorithm systematically explores the entire parameter space. Thus, it will encounter the infamous curse of dimensionality [Sch11] when dealing with high-dimensional issues.

## 6.4 Future Directions

### 6.4.1 Sensor Extrinsic Calibration

**Automatic and online calibration.** The proposed extrinsic calibration method is only suitable for offline calibration since it needs several manual operations for data acquisition and preprocessing. The next step should be to make the entire calibration process fully automatic and online, enabling calibration in various situations to ensure the perceptual stability of the autonomous system. The measurements from another roadside sensor, such as the camera, can be used to enable data association with the roadside radar. For instance, multi-object tracking technology can be employed to extract and match traffic objects.

On the other hand, the proposed globally optimal registration method in the back-end is robust, but also time-consuming. The current optimization back-end should be more efficient while ensuring robustness to outliers and noise. A possible direction is utilizing GPU

and parallel computing to accelerate the BnB algorithm. From the optimization perspective, exploring a new optimization paradigm can also avoid computational limitations.

**Spatio-temporal calibration.** The temporal calibration between heterogeneous sensors is also crucial, which can ensure reliable long-term operation of sensors. Heterogeneous sensors commonly have various sampling rates. Thus, they have an inherent temporal misalignment. The proposed calibration method considers this time synchronization problem. However, the temporal parameters are not integrated into the optimization process. It is valuable to simultaneously optimize the spatial and temporal parameters to achieve accurate time drift and extrinsic parameter estimation.

**Robust calibration for multi-sensors.** To ensure a reliable environment understanding, most perception systems incorporate multiple sensors and sensing modalities, such as LiDAR, camera, and event camera. In practice, the performance of cross-modal feature extraction is brittle, which may result in severe noise and outliers. As illustrated in the proposed calibration method, the BnB algorithm achieves significant performance in robust parameter estimation. Therefore, introducing BnB algorithms in the extrinsic calibration of multi-modal sensors is a challenging task but it has great significance in enhancing the robustness against noisy and outlier-contaminated input.

## 6.4.2 Point Cloud Registration

**Robust registration with optimality certification.** Although the proposed point cloud registration approaches are highly efficient and robust, a shared and inherent drawback is that they cannot theoretically guarantee global optimality in terms of inlier set maximization. The transformation decoupling strategy relaxes the original constraint geometrically so that the decoupled sub-problems are not exactly the same as the raw registration problem. Hence, one potential direction is to explore robust registration methods with optimality certification [YC22; Ant+21]. Such methods should be capable of verifying the optimality of the obtained estimation results and, importantly, have the ability to identify cases of failed estimation rather than blindly returning incorrect results.

**Novel transformation decoupling strategy.** This thesis presents three different transformation decoupling strategies according to different geometric properties and achieves promising registration results. This encourages us to deeply investigate the application of transformation decoupling strategies in geometry perception. For instance, researchers commonly address the rotation search problem, a vital part of pose estimation, by BnB algorithms. Developing the rotation decoupling strategy probably facilitates a good trade-off between efficiency and robustness, while promoting the BnB algorithm a step further in practical applications. In addition, we initially explore how to accelerate BnB with the interval stabbing technique in Chapter 4, which allows us to branch over a two-dimensional space for the three-dimensional sub-problem. It would be interesting to further explore the application of

this technique to other geometric problems for accelerating the BnB algorithm, such as the relative pose estimation problem.

**Pose estimation assisted with prior information.** The proposed methods in Chapter 5 utilize the prior known gravity directions to assist point cloud registration, thus significantly increasing the registration efficiency. This is a good demonstration of utilizing prior information or reasonable assumptions to help pose estimation. Therefore, leveraging additional prior information or assumptions to assist a broader range of pose estimation applications may offer potential benefits. The prior information or assumptions that can possibly be utilized include the structural world assumption (e.g., Manhattan and Atlanta worlds), geometric constraints (e.g., planar surfaces, intersecting lines, or parallel lines), semantic information, motion model, etc.

**Robust solutions for other pose estimation problems.** This thesis addresses two typical pose estimation problems, including sensor extrinsic calibration and point cloud registration. However, pose estimation technology still has many other applications, such as camera absolute pose estimation, relative pose estimation, vehicle localization, and direction estimation. These problems have different geometric properties and constraints. For instance, both the direction estimation and the relative pose estimation problem have the unit-norm constraint. Besides, absolute pose estimation (i.e., 2D-3D registration) is inherently more difficult than point cloud registration. Therefore, it is challenging but valuable to explore geometric properties and develop novel robust methods to address these problems.



## Bibliography

- [AMC08] Aiger, D., Mitra, N. J., and Cohen-Or, D. “4-points congruent sets for robust pairwise surface registration”. In: *ACM SIGGRAPH 2008 papers*. 2008, pp. 1–10.
- [ART10] Albarelli, A., Rodola, E., and Torsello, A. “A game-theoretic approach to fine surface registration without initial motion estimation”. In: *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE. 2010, pp. 430–437.
- [ADV03] Almansa, A., Desolneux, A., and Vamech, S. “Vanishing point detection without any a priori information”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25.4 (2003), pp. 502–507.
- [Ant+21] Antonante, P., Tzoumas, V., Yang, H., and Carlone, L. “Outlier-robust estimation: Hardness, minimally tuned algorithms, and applications”. In: *IEEE Transactions on Robotics* 38.1 (2021), pp. 281–301.
- [Ao+21] Ao, S., Hu, Q., Yang, B., Markham, A., and Guo, Y. “Spinnet: Learning a general surface descriptor for 3d point cloud registration”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pp. 11753–11762.
- [Aok+19] Aoki, Y., Goforth, H., Srivatsan, R. A., and Lucey, S. “Pointnetlk: Robust & efficient point cloud registration using pointnet”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 7163–7172.
- [Arn+20] Arnold, E., Dianati, M., Temple, R. de, and Fallah, S. “Cooperative perception for 3D object detection in driving scenarios using infrastructure sensors”. In: *IEEE Transactions on Intelligent Transportation Systems* (2020).
- [AHB87] Arun, K. S., Huang, T. S., and Blostein, S. D. “Least-squares fitting of two 3-D point sets”. In: *IEEE Transactions on pattern analysis and machine intelligence* 5 (1987), pp. 698–700.
- [Bai+21] Bai, X., Luo, Z., Zhou, L., Chen, H., Li, L., Hu, Z., Fu, H., and Tai, C.-L. “Pointdsc: Robust point cloud registration using deep spatial consistency”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 15859–15869.

- [Bai+20] Bai, X., Luo, Z., Zhou, L., Fu, H., Quan, L., and Tai, C.-L. “D3feat: Joint learning of dense detection and description of 3d local features”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 6359–6367.
- [BJX20] Bai, Z., Jiang, G., and Xu, A. “Lidar-camera calibration using line correspondences”. In: *Sensors* 20.21 (2020), p. 6319.
- [Bal76] Ball, R. S. “The theory of screws: A study in the dynamics of a rigid body”. In: *Mathematische Annalen* 9.4 (1876), pp. 541–553.
- [Bar+21] Barath, D., Ding, Y., Kukulova, Z., and Larsson, V. “Image Stitching with Locally Shared Rotation Axis”. In: *2021 International Conference on 3D Vision (3DV)*. IEEE. 2021, pp. 1382–1391.
- [BM21] Barath, D. and Matas, J. “Graph-cut RANSAC: Local optimization on spatially coherent structures”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.9 (2021), pp. 4961–4974.
- [BNM21] Barath, D., Noskova, J., and Matas, J. “Marginalizing sample consensus”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.11 (2021), pp. 8420–8432.
- [BSP12] Bazin, J.-C., Seo, Y., and Pollefeys, M. “Globally optimal consensus set maximization through rotation search”. In: *Asian Conference on Computer Vision*. Springer. 2012, pp. 539–551.
- [Bel+22] Beltrán, J., Guindel, C., Escalera, A. de la, and Garcia, F. “Automatic extrinsic calibration method for lidar and camera sensor setups”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.10 (2022), pp. 17677–17689.
- [BM92] Besl, P. and McKay, N. D. “A method for registration of 3-D shapes”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2 (1992), pp. 239–256.
- [BS03] Biber, P. and Straßer, W. “The normal distributions transform: A new approach to laser scan matching”. In: *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*. Vol. 3. IEEE. 2003, pp. 2743–2748.
- [BL95] Blais, G. and Levine, M. D. “Registering multiview range data to create 3D computer objects”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17.8 (1995), pp. 820–824.
- [BEN13] Borrmann, D., Elseberg, J., and Nüchter, A. “Thermal 3D mapping of building façades”. In: *Intelligent Autonomous Systems 12: Volume 1 Proceedings of the 12th International Conference IAS-12, held June 26-29, 2012, Jeju Island, Korea*. Springer. 2013, pp. 173–182.
- [BR90] Bottema, O. and Roth, B. *Theoretical kinematics*. Vol. 24. Courier Corporation, 1990.



- [BTP13] Bouaziz, S., Tagliasacchi, A., and Pauly, M. “Sparse iterative closest point”. In: *Computer graphics forum*. Vol. 32. 5. Wiley Online Library. 2013, pp. 113–123.
- [BWG19] Brown, M., Windridge, D., and Guillemaut, J.-Y. “A family of globally optimal branch-and-bound algorithms for 2D–3D correspondence-free registration”. In: *Pattern Recognition 93* (2019), pp. 36–54.
- [Bus+19] Bustos, A. P., Chin, T.-J., Neumann, F., Friedrich, T., and Katzmann, M. “A practical maximum clique algorithm for matching with pairwise constraints”. In: *arXiv preprint arXiv:1902.01534* (2019).
- [Bus+16] Bustos, A. P., Chin, T., Eriksson, A., Li, H., and Suter, D. “Fast Rotation Search with Stereographic Projections for 3D Registration”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38.11 (2016), pp. 2227–2240. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2016.2517636.
- [BC17] Bustos, A. P. and Chin, T.-J. “Guaranteed outlier removal for point cloud registration with correspondences”. In: *IEEE transactions on pattern analysis and machine intelligence* 40.12 (2017), pp. 2868–2882.
- [Cai+19] Cai, Z., Chin, T.-J., Bustos, A. P., and Schindler, K. “Practical optimal registration of terrestrial LiDAR scan pairs”. In: *ISPRS journal of photogrammetry and remote sensing* 147 (2019), pp. 118–131.
- [CCK19] Cai, Z., Chin, T.-J., and Koltun, V. “Consensus maximization tree search revisited”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 1637–1645.
- [CP15] Campbell, D. and Petersson, L. “An adaptive data representation for robust point-set registration and merging”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 4292–4300.
- [CP16] Campbell, D. and Petersson, L. “Gogma: Globally-optimal gaussian mixture alignment”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 5685–5694.
- [Cam+18] Campbell, D., Petersson, L., Kneip, L., and Li, H. “Globally-optimal inlier set maximisation for camera pose and correspondence estimation”. In: *IEEE transactions on pattern analysis and machine intelligence* 42.2 (2018), pp. 328–342.
- [Cam+19] Campbell, D., Petersson, L., Kneip, L., Li, H., and Gould, S. “The alignment of the spheres: Globally-optimal spherical mixture alignment for camera pose estimation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 11796–11806.
- [CVV22] Cattaneo, D., Vaghi, M., and Valada, A. “Lcdnet: Deep loop closure detection and point cloud registration for lidar slam”. In: *IEEE Transactions on Robotics* 38.4 (2022), pp. 2074–2093.

- [CA15] Chavez-Garcia, R. O. and Aycard, O. “Multiple sensor fusion and classification for moving object detection and tracking”. In: *IEEE Transactions on Intelligent Transportation Systems* 17.2 (2015), pp. 525–534.
- [Che+22a] Chen, F., Li, L., Zhang, S., Wu, J., and Wang, L. “PBACalib: Targetless Extrinsic Calibration for High-Resolution LiDAR-Camera System Based on Plane-Constrained Bundle Adjustment”. In: *IEEE Robotics and Automation Letters* 8.1 (2022), pp. 304–311.
- [Che+20] Chen, G., Cao, H., Conradt, J., Tang, H., Rohrbein, F., and Knoll, A. “Event-Based Neuromorphic Vision for Autonomous Driving: A Paradigm Shift for Bio-Inspired Visual Sensing and Perception”. In: *IEEE Signal Processing Magazine* 37.4 (2020), pp. 34–49. DOI: 10.1109/MSP.2020.2985815.
- [Che+22b] Chen, H., Wang, P., Wang, F., Tian, W., Xiong, L., and Li, H. “Epro-pnp: Generalized end-to-end probabilistic perspective-n-points for monocular object pose estimation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 2781–2790.
- [Che+09] Chen, L., Guo, L., Yang, N., and Du, Y. “Multi-Level Image Thresholding Based on Histogram Voting”. In: *2009 2nd International Congress on Image and Signal Processing*. IEEE. 2009, pp. 1–5.
- [Che+22c] Chen, W., Li, H., Nie, Q., and Liu, Y.-H. “Deterministic Point Cloud Registration via Novel Transformation Decomposition”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 6348–6356.
- [Che+23a] Chen, Y., Liu, J., Zhang, X., Qi, X., and Jia, J. “Voxelnext: Fully sparse voxelnet for 3d object detection and tracking”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 21674–21683.
- [Che+23b] Chen, Z., Sun, K., Yang, F., Guo, L., and Tao, W. “SC2-PCR++: Rethinking the Generation and Selection for Efficient and Robust Point Cloud Registration”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).
- [Che+22d] Chen, Z., Sun, K., Yang, F., and Tao, W. “SC2-PCR: A Second Order Spatial Compatibility for Efficient and Robust Point Cloud Registration”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 13221–13231.
- [CYT22] Chen, Z., Yang, F., and Tao, W. “Detarnet: Decoupling translation and rotation by siamese network for point cloud registration”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. 1. 2022, pp. 401–409.
- [CSK05] Chetverikov, D., Stepanov, D., and Krsek, P. “Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm”. In: *Image and vision computing* 23.3 (2005), pp. 299–309.

- [Chi+17] Chin, T.-J., Purkait, P., Eriksson, A., and Suter, D. “Efficient Globally Optimal Consensus Maximisation with Tree Search”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.4 (2017), pp. 758–772. DOI: 10.1109/TPAMI.2016.2631531.
- [CS17] Chin, T.-J. and Suter, D. “The maximum consensus problem: recent algorithmic advances”. In: *Synthesis Lectures on Computer Vision* 7.2 (2017), pp. 1–194.
- [CYS17] Chou, C., Yeh, S.-H., and Song, D. “Mirror-assisted calibration of a multi-modal sensing array with a ground penetrating radar and a camera”. In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2017, pp. 1457–1463.
- [CDK20] Choy, C., Dong, W., and Koltun, V. “Deep global registration”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 2514–2523.
- [CPK19] Choy, C., Park, J., and Koltun, V. “Fully convolutional geometric features”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 8958–8966.
- [CMK03] Chum, O., Matas, J., and Kittler, J. “Locally optimized RANSAC”. In: *Pattern Recognition: 25th DAGM Symposium, Magdeburg, Germany, September 10-12, 2003. Proceedings 25*. Springer. 2003, pp. 236–243.
- [Cla99] Clausen, J. “Branch and bound algorithms-principles and examples”. In: *Department of Computer Science, University of Copenhagen* (1999), pp. 1–30.
- [Cre+22] Creß, C., Zimmer, W., Strand, L., Fortkord, M., Dai, S., Lakshminarasimhan, V., and Knoll, A. “A9-dataset: Multi-sensor infrastructure-based dataset for mobility research”. In: *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2022, pp. 965–970.
- [CL96] Curless, B. and Levoy, M. “A volumetric method for building complex models from range images”. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. 1996, pp. 303–312.
- [Das+22] Das, S., Mahabadi, N., Djikic, A., Nassir, C., Chatterjee, S., and Fallon, M. “Extrinsic calibration and verification of multiple non-overlapping field of view lidar sensors”. In: *2022 International Conference on Robotics and Automation (ICRA)*. IEEE. 2022, pp. 919–925.
- [De +97] De Berg, M., Van Kreveld, M., Overmars, M., Schwarzkopf, O., Berg, M. de, Kreveld, M. van, Overmars, M., and Schwarzkopf, O. *Computational Geometry: Introduction*. Springer, 1997.
- [DBK21] Ding, Y., Barath, D., and Kukulova, Z. “Minimal solutions for panoramic stitching given gravity prior”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 5579–5588.

- [Din+21] Ding, Y., Barath, D., Yang, J., Kong, H., and Kukulova, Z. “Globally optimal relative pose estimation with gravity prior”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 394–403.
- [Din+20] Ding, Y., Yang, J., Ponce, J., and Kong, H. “Homography-based minimal-case relative pose estimation with known gravity direction”. In: *IEEE transactions on pattern analysis and machine intelligence* 44.1 (2020), pp. 196–210.
- [DKG21] Domhof, J., Kooij, J. F., and Gavrila, D. M. “A Joint Extrinsic Calibration Tool for Radar, Camera and Lidar”. In: *IEEE Transactions on Intelligent Vehicles* (2021).
- [Don+20] Dong, Z., Liang, F., Yang, B., Xu, Y., Zang, Y., Li, J., Wang, Y., Dai, W., Fan, H., Hyyppä, J., et al. “Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 163 (2020), pp. 327–342.
- [Du+21] Du, Y., Qin, B., Zhao, C., Zhu, Y., Cao, J., and Ji, Y. “A novel spatio-temporal synchronization method of roadside asynchronous MMW radar-camera for sensor fusion”. In: *IEEE Transactions on Intelligent Transportation Systems* (2021).
- [Dub+14] Dubská, M., Herout, A., Juránek, R., and Sochor, J. “Fully automatic roadside camera calibration for traffic surveillance”. In: *IEEE Transactions on Intelligent Transportation Systems* 16.3 (2014), pp. 1162–1171.
- [El +15] El Natour, G., Aider, O. A., Rouveure, R., Berry, F., and Faure, P. “Radar and vision sensors calibration for outdoor 3D reconstruction”. In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2015, pp. 2084–2089.
- [Fan+21] Fan, A., Ma, J., Jiang, X., and Ling, H. “Efficient deterministic search with robust loss functions for geometric model fitting”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.11 (2021), pp. 8212–8229.
- [FB81] Fischler, M. A. and Bolles, R. C. “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography”. In: *Communications of the ACM* 24.6 (1981), pp. 381–395.
- [Fit03] Fitzgibbon, A. W. “Robust registration of 2D and 3D point sets”. In: *Image and vision computing* 21.13-14 (2003), pp. 1145–1153.
- [Flo13] Floudas, C. A. *Deterministic global optimization: theory, methods and applications*. Vol. 37. Springer Science & Business Media, 2013.
- [Fu+21] Fu, K., Liu, S., Luo, X., and Wang, M. “Robust point cloud registration framework based on deep graph matching”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 8893–8902.
- [Ge17] Ge, X. “Automatic markerless registration of point clouds with semantic-keypoint-based 4-points congruent sets”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 130 (2017), pp. 344–357.

- [GLU12] Geiger, A., Lenz, P., and Urtasun, R. “Are we ready for autonomous driving? the kitti vision benchmark suite”. In: *2012 IEEE conference on computer vision and pattern recognition*. IEEE. 2012, pp. 3354–3361.
- [Gei+12] Geiger, A., Moosmann, F., Car, Ö., and Schuster, B. “Automatic camera and range sensor calibration using a single shot”. In: *2012 IEEE international conference on robotics and automation*. IEEE. 2012, pp. 3936–3943.
- [Gle+14] Glent Buch, A., Yang, Y., Kruger, N., and Gordon Petersen, H. “In search of inliers: 3d correspondence by local and global voting”. In: *Proceedings of the IEEE Conference on computer vision and pattern recognition*. 2014, pp. 2067–2074.
- [Goj+19] Gojcic, Z., Zhou, C., Wegner, J. D., and Wieser, A. “The perfect match: 3d point cloud matching with smoothed densities”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 5545–5554.
- [GRG09] González-Aguilera, D., Rodríguez-Gonzálvez, P., and Gómez-Lahoz, J. “An automatic procedure for co-registration of terrestrial laser scanners and digital cameras”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 64.3 (2009), pp. 308–316.
- [GP02] Granger, S. and Pennec, X. “Multi-scale EM-ICP: A fast and robust approach for surface registration”. In: *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part IV* 7. Springer. 2002, pp. 418–432.
- [Guo+14] Guo, Y., Bennamoun, M., Sohel, F., Lu, M., and Wan, J. “3D object recognition in cluttered scenes with local surface features: A survey”. In: *IEEE transactions on pattern analysis and machine intelligence* 36.11 (2014), pp. 2270–2287.
- [Guo+16] Guo, Y., Bennamoun, M., Sohel, F., Lu, M., Wan, J., and Kwok, N. M. “A comprehensive performance evaluation of 3D local feature descriptors”. In: *International Journal of Computer Vision* 116.1 (2016), pp. 66–89.
- [Guo+13a] Guo, Y., Sohel, F., Bennamoun, M., Lu, M., and Wan, J. “Rotational projection statistics for 3D local surface description and object recognition”. In: *International journal of computer vision* 105 (2013), pp. 63–86.
- [Guo+13b] Guo, Y., Sohel, F., Bennamoun, M., Lu, M., and Wan, J. “TriSI: A distinctive local surface descriptor for 3D modeling and object recognition”. In: *International Conference on Computer Graphics Theory and Applications*. Vol. 2. Scitepress. 2013, pp. 86–93.
- [Guo+20] Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., and Bennamoun, M. “Deep learning for 3d point clouds: A survey”. In: *IEEE transactions on pattern analysis and machine intelligence* 43.12 (2020), pp. 4338–4364.
- [HK09] Hartley, R. I. and Kahl, F. “Global optimization through rotation space search”. In: *International Journal of Computer Vision* 82.1 (2009), pp. 64–79.

- [Hen20] Heng, L. “Automatic Targetless Extrinsic Calibration of Multiple 3D LiDARs and Radars”. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020, pp. 10669–10675.
- [Hir21] Hirose, O. “A Bayesian Formulation of Coherent Point Drift”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.7 (2021), pp. 2269–2286. DOI: 10.1109/TPAMI.2020.2971687.
- [HK17] Horanyi, N. and Kato, Z. “Multiview absolute pose using 3D-2D perspective line correspondences and vertical direction”. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017, pp. 2472–2480.
- [Hor87] Horn, B. K. “Closed-form solution of absolute orientation using unit quaternions”. In: *Josa a* 4.4 (1987), pp. 629–642.
- [Hua+21] Huang, S., Gojcic, Z., Usvyatsov, M., Wieser, A., and Schindler, K. “Predator: Registration of 3d point clouds with low overlap”. In: *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*. 2021, pp. 4267–4276.
- [HH78] Hunt, K. H. and Hunt, K. H. *Kinematic geometry of mechanisms*. Vol. 7. Oxford University Press, USA, 1978.
- [IO118] Ishikawa, R., Oishi, T., and Ikeuchi, K. “Lidar and camera calibration using motions estimated by sensor fusion odometry”. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2018, pp. 7342–7349.
- [Izq+18] Izquierdo, R., Parra, I., Fernández-Llorca, D., and Sotelo, M. “Multi-radar self-calibration method using high-definition digital maps for autonomous driving”. In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2018, pp. 2197–2202.
- [JV05] Jian, B. and Vemuri, B. C. “A robust algorithm for point set registration using mixture of Gaussians”. In: *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*. Vol. 2. IEEE. 2005, pp. 1246–1251.
- [JV10] Jian, B. and Vemuri, B. C. “Robust point set registration using gaussian mixture models”. In: *IEEE transactions on pattern analysis and machine intelligence* 33.8 (2010), pp. 1633–1645.
- [Jia+23a] Jiang, H., Dang, Z., Wei, Z., Xie, J., Yang, J., and Salzmann, M. “Robust Outlier Rejection for 3D Registration with Variational Bayes”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 1148–1157.
- [Jia+23b] Jiao, J., Chen, F., Wei, H., Wu, J., and Liu, M. “LCE-Calib: Automatic LiDAR-Frame/Event Camera Extrinsic Calibration With a Globally Optimal Solution”. In: *IEEE/ASME Transactions on Mechatronics* (2023), pp. 1–12. DOI: 10.1109/TMECH.2023.3259444.

- [Jia+21] Jiao, J., Ye, H., Zhu, Y., and Liu, M. “Robust odometry and mapping for multi-lidar systems with online extrinsic calibration”. In: *IEEE Transactions on Robotics* 38.1 (2021), pp. 351–371.
- [Jia+22] Jiao, Y., Wang, Y., Ding, X., Wang, M., and Xiong, R. “Deterministic Optimality for Robust Vehicle Localization Using Visual Measurements”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.6 (2022), pp. 5397–5410. DOI: 10.1109/TITS.2021.3053574.
- [KH08] Kahl, F. and Hartley, R. “Multiple-View Geometry Under the  $L_\infty$ -Norm”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.9 (2008), pp. 1603–1617. DOI: 10.1109/TPAMI.2007.70824.
- [KD20] Kang, J. and Doh, N. L. “Automatic targetless camera–lidar calibration by aligning edge with gaussian mixture model”. In: *Journal of Field Robotics* 37.1 (2020), pp. 158–179.
- [KN10] Khan, M. K. and Nyström, I. “A modified particle swarm optimization applied in image registration”. In: *2010 20th International Conference on Pattern Recognition*. IEEE. 2010, pp. 2302–2305.
- [Krä+19] Krämmer, A., Schöllner, C., Gulati, D., Lakshminarasimhan, V., Kurz, F., Rosenbaum, D., Lenz, C., and Knoll, A. “Providentia—a large-scale sensor system for the assistance of autonomous vehicles and its evaluation”. In: *arXiv preprint arXiv:1906.06789* (2019).
- [KVP22] Kubelka, V., Vaidis, M., and Pomerleau, F. “Gravity-constrained point cloud registration”. In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2022, pp. 4873–4879. DOI: 10.1109/IROS47612.2022.9981916.
- [KK20] Kümmerle, J. and Kühner, T. “Unified intrinsic and extrinsic camera and LiDAR calibration under uncertainties”. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2020, pp. 6028–6034.
- [KKL18] Kümmerle, J., Kühner, T., and Lauer, M. “Automatic calibration of multiple cameras and depth sensors with a spherical target”. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2018, pp. 1–8.
- [Le+19] Le, H., Chin, T.-J., Eriksson, A., Do, T.-T., and Suter, D. “Deterministic approximate methods for maximum consensus robust fitting”. In: *IEEE transactions on pattern analysis and machine intelligence* 43.3 (2019), pp. 842–857.
- [LMC12] Lebeda, K., Matas, J., and Chum, O. “Fixing the locally optimized ransac–full experimental evaluation”. In: *British machine vision conference*. Vol. 2. Citeseer. 2012.
- [Lec+19] Lecrosnier, L., Bouteau, R., Vasseur, P., Savatier, X., and Fraundorfer, F. “Camera pose estimation based on PnL with a known vertical direction”. In: *IEEE Robotics and Automation Letters* 4.4 (2019), pp. 3852–3859.

- [Lee+20] Lee, C.-L., Hsueh, Y.-H., Wang, C.-C., and Lin, W.-C. “Extrinsic and temporal calibration of automotive radar and 3D LiDAR”. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020, pp. 9976–9983.
- [Lee+21] Lee, J., Kim, S., Cho, M., and Park, J. “Deep hough voting for robust global registration”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 15994–16003.
- [LT13] Levinson, J. and Thrun, S. “Automatic Online Calibration of Cameras and Lasers”. In: *Robotics: science and systems*. Vol. 2. 7. Berlin, Germany. 2013.
- [Li+23a] Li, H., Zhao, J., Bazin, J.-C., Kim, P., Joo, K., Zhao, Z., and Liu, Y.-H. “Hong Kong World: Leveraging Structural Regularity for Line-based SLAM”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023), pp. 1–18. DOI: 10.1109/TPAMI.2023.3276204.
- [Li09] Li, H. “Consensus set maximization with guaranteed global optimality for robust geometry estimation”. In: *2009 IEEE 12th International Conference on Computer Vision*. 2009, pp. 1074–1080.
- [LH07] Li, H. and Hartley, R. “The 3D-3D registration problem revisited”. In: *2007 IEEE 11th international conference on computer vision*. IEEE. 2007, pp. 1–8.
- [LL19] Li, J. and Lee, G. H. “Usip: Unsupervised stable interest point detection from 3d point clouds”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 361–370.
- [Li22] Li, J. “A Practical  $O(N^2)$  Outlier Removal Method for Correspondence-Based Point Cloud Registration”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.8 (2022), pp. 3926–3939. DOI: 10.1109/TPAMI.2021.3065021.
- [LHA20] Li, J., Hu, Q., and Ai, M. “GESAC: Robust graph enhanced sample consensus for point cloud registration”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 167 (2020), pp. 363–374.
- [LHA21] Li, J., Hu, Q., and Ai, M. “Point cloud registration based on one-point ransac and scale-annealing biweight estimation”. In: *IEEE Transactions on Geoscience and Remote Sensing* 59.11 (2021), pp. 9716–9729.
- [Li+22] Li, J., Hu, Q., Zhang, Y., and Ai, M. “Robust symmetric iterative closest point”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 185 (2022), pp. 219–231.
- [Li+23b] Li, J., Shi, P., Hu, Q., and Zhang, Y. “QGORE: Quadratic-Time Guaranteed Outlier Removal for Point Cloud Registration”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023), pp. 1–16. DOI: 10.1109/TPAMI.2023.3262780.



- [Li+20] Li, J., Zhao, P., Hu, Q., and Ai, M. “Robust point cloud registration based on topological graph and Cauchy weighted lq-norm”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 160 (2020), pp. 244–259.
- [Li+17] Li, L., Yang, M., Wang, C., and Wang, B. “Rigid point set registration based on cubature Kalman filter and its application in intelligent vehicles”. In: *IEEE Transactions on Intelligent Transportation Systems* 19.6 (2017), pp. 1754–1765.
- [Li+23c] Li, X., Liu, Y., Cao, H., Liu, X., Zhang, F., and Knoll, A. “Efficient and Deterministic Search Strategy Based on Residual Projections for Point Cloud Registration”. In: *arXiv preprint arXiv:2305.11716* (2023).
- [Li+23d] Li, X., Liu, Y., Lakshminarasimhan, V., Cao, H., Zhang, F., and Knoll, A. “Globally Optimal Robust Radar Calibration in Intelligent Transportation Systems”. In: *IEEE Transactions on Intelligent Transportation Systems* 24.6 (2023), pp. 6082–6095. DOI: 10.1109/TITS.2023.3251183.
- [Li+23e] Li, X., Liu, Y., Xia, Y., Lakshminarasimhan, V., Cao, H., Zhang, F., Stilla, U., and Knoll, A. “Fast and deterministic (3+ 1) DOF point set registration with gravity prior”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 199 (2023), pp. 118–132.
- [Li+23f] Li, X., Ma, Z., Liu, Y., Zimmer, W., Cao, H., Zhang, F., and Knoll, A. “Transformation Decoupling Strategy based on Screw Theory for Deterministic Point Cloud Registration with Gravity Prior”. In: *arXiv preprint arXiv:2311.01432* (2023).
- [Li+18] Li, X., Liu, Y., Wang, Y., Wang, C., Wang, M., and Song, Z. “Fast and globally optimal rigid registration of 3d point sets by transformation decomposition”. In: *arXiv preprint arXiv:1812.11307* (2018).
- [Li+23g] Li, Y., Yin, Z., Zheng, Y., Lu, H., Kamiya, T., Nakatoh, Y., and Serikawa, S. “Pose Estimation of Point Sets Using Residual MLP in Intelligent Transportation Infrastructure”. In: *IEEE Transactions on Intelligent Transportation Systems* (2023), pp. 1–11. DOI: 10.1109/TITS.2023.3250604.
- [LZY16] Lian, W., Zhang, L., and Yang, M.-H. “An efficient globally optimal algorithm for asymmetric point matching”. In: *IEEE transactions on pattern analysis and machine intelligence* 39.7 (2016), pp. 1281–1293.
- [Lia+23] Liao, Y., Li, J., Kang, S., Li, Q., Zhu, G., Yuan, S., Dong, Z., and Yang, B. “SE-Calib: Semantic Edges based LiDAR-Camera Bore-sight Online Calibration in Urban Scenes”. In: *IEEE Transactions on Geoscience and Remote Sensing* (2023).
- [Lim+22] Lim, H., Yeon, S., Ryu, S., Lee, Y., Kim, Y., Yun, J., Jung, E., Lee, D., and Myung, H. “A Single Correspondence Is Enough: Robust Global Registration to Avoid Degeneracy in Urban Environments”. In: *2022 International Conference on Robotics and Automation (ICRA)*. 2022, pp. 8010–8017. DOI: 10.1109/ICRA46639.2022.9812018.

- [LMD21] Lim, T.-Y., Markowitz, S. A., and Do, M. N. “RaDICaL: A Synchronized FMCW Radar, Depth, IMU and RGB Camera Data Dataset With Low-Level FMCW Radar Signals”. In: *IEEE Journal of Selected Topics in Signal Processing* 15.4 (2021), pp. 941–953.
- [LWC21] Liu, W., Wu, H., and Chirikjian, G. S. “LSG-CPD: Coherent point drift with local surface geometry for point cloud registration”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 15293–15302.
- [LYZ22] Liu, X., Yuan, C., and Zhang, F. “Targetless extrinsic calibration of multiple small FoV LiDARs and cameras using adaptive voxelization”. In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), pp. 1–12.
- [LZ21] Liu, X. and Zhang, F. “Extrinsic calibration of multiple lidars of small fov in targetless environments”. In: *IEEE Robotics and Automation Letters* 6.2 (2021), pp. 2036–2043.
- [Liu+21] Liu, Y., Chen, G., Gu, R., and Knoll, A. “Globally Optimal Consensus Maximization for Relative Pose Estimation With Known Gravity Direction”. In: *IEEE Robotics and Automation Letters* 6.3 (2021), pp. 5905–5912.
- [LCK20a] Liu, Y., Chen, G., and Knoll, A. “Globally optimal camera orientation estimation from line correspondences by bnb algorithm”. In: *IEEE Robotics and Automation Letters* 6.1 (2020), pp. 215–222.
- [LCK20b] Liu, Y., Chen, G., and Knoll, A. “Globally optimal vertical direction estimation in atlanta world”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.4 (2020), pp. 1949–1962.
- [LCK23] Liu, Y., Chen, G., and Knoll, A. “Absolute Pose Estimation with a Known Direction by Motion Decoupling”. In: *IEEE Transactions on Circuits and Systems for Video Technology* (2023).
- [Liu+18a] Liu, Y., Dong, Y., Song, Z., and Wang, M. “2d-3d point set registration based on global rotation search”. In: *IEEE Transactions on Image Processing* 28.5 (2018), pp. 2599–2613.
- [Liu+18b] Liu, Y., Wang, C., Song, Z., and Wang, M. “Efficient global point cloud registration by matching rotation invariant features through translation search”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 448–463.
- [Liu+22] Liu, Y., Wang, Y., Wang, M., Chen, G., Knoll, A., and Song, Z. “Globally optimal linear model fitting with unit-norm constraint”. In: *International Journal of Computer Vision* 130.4 (2022), pp. 933–946.
- [Low04] Lowe, D. G. “Distinctive image features from scale-invariant keypoints”. In: *International journal of computer vision* 60 (2004), pp. 91–110.

- [Lu+20] Lu, F., Chen, G., Liu, Y., Qu, Z., and Knoll, A. “Rskdd-net: Random sample-based keypoint detector and descriptor”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 21297–21308.
- [LLH00] Luck, J., Little, C., and Hoff, W. “Registration of range data using a hybrid simulated annealing and iterative closest point algorithm”. In: *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*. Vol. 4. IEEE. 2000, pp. 3739–3744.
- [LP17] Lynch, K. M. and Park, F. C. *Modern robotics*. Cambridge University Press, 2017.
- [Mai+11] Maier-Hein, L., Franz, A. M., Dos Santos, T. R., Schmidt, M., Fangerau, M., Meinzer, H.-P., and Fitzpatrick, J. M. “Convergent iterative closest-point algorithm to accomodate anisotropic and inhomogenous localization error”. In: *IEEE transactions on pattern analysis and machine intelligence* 34.8 (2011), pp. 1520–1532.
- [MSS08] Mehlhorn, K., Sanders, P., and Sanders, P. *Algorithms and data structures: The basic toolbox*. Vol. 55. Springer, 2008.
- [MAM14] Mellado, N., Aiger, D., and Mitra, N. J. “Super 4pcs fast global pointcloud registration via smart indexing”. In: *Computer graphics forum*. Vol. 33. 5. Wiley Online Library. 2014, pp. 205–215.
- [Men+17] Menouar, H., Guvenc, I., Akkaya, K., Uluagac, A. S., Kadri, A., and Tuncer, A. “UAV-enabled intelligent transportation systems for the smart city: Applications and challenges”. In: *IEEE Communications Magazine* 55.3 (2017), pp. 22–28.
- [MBO06] Mian, A. S., Bennamoun, M., and Owens, R. A. “A novel representation and feature matching algorithm for automatic pairwise registration of range images”. In: *International Journal of Computer Vision* 66 (2006), pp. 19–40.
- [Mis+20] Mishra, S., Osteen, P. R., Pandey, G., and Saripalli, S. “Experimental evaluation of 3d-lidar camera extrinsic calibration”. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020, pp. 9020–9026.
- [MBZ13] Moghadam, P., Bosse, M., and Zlot, R. “Line-based extrinsic calibration of range and image sensors”. In: *2013 IEEE International Conference on Robotics and Automation*. IEEE. 2013, pp. 3685–3691.
- [Mor+16] Morrison, D. R., Jacobson, S. H., Sauppe, J. J., and Sewell, E. C. “Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning”. In: *Discrete Optimization* 19 (2016), pp. 79–102.
- [MS10] Myronenko, A. and Song, X. “Point set registration: Coherent point drift”. In: *IEEE transactions on pattern analysis and machine intelligence* 32.12 (2010), pp. 2262–2275.
- [Nes+18] Nesterov, Y. et al. *Lectures on convex optimization*. Vol. 137. Springer, 2018.

- [Oh+18] Oh, J., Kim, K.-S., Park, M., and Kim, S. “A comparative study on camera-radar calibration methods”. In: *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE. 2018, pp. 1057–1062.
- [OKO08] Olsson, C., Kahl, F., and Oskarsson, M. “Branch-and-bound methods for euclidean registration problems”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31.5 (2008), pp. 783–794.
- [Olu+21] Olutomilayo, K. T., Bahramgiri, M., Nooshabadi, S., and Fuhrmann, D. R. “Extrinsic calibration of radar Mount position and orientation with multiple target configurations”. In: *IEEE Transactions on Instrumentation and Measurement* 70 (2021), pp. 1–13.
- [Örn+22] Örnå, M. V., Persson, P., Wadenbäck, M., Åström, K., and Heyden, A. “Trust Your IMU: Consequences of Ignoring the IMU Drift”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 4468–4477.
- [Pai+20] Pais, G. D., Ramalingam, S., Govindu, V. M., Nascimento, J. C., Chellappa, R., and Miraldo, P. “3dregnet: A deep neural network for 3d point registration”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 7193–7203.
- [Pav+18] Pavlov, A. L., Ovchinnikov, G. W., Derbyshev, D. Y., Tsetserukou, D., and Osleledets, I. V. “AA-ICP: Iterative closest point with Anderson acceleration”. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2018, pp. 3407–3412.
- [PTV22] Peng, L., Tsakiris, M. C., and Vidal, R. “ARCS: Accurate Rotation and Correspondence Search”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 11153–11163.
- [Pen+20] Pentek, Q., Kennel, P., Allouis, T., Fiorio, C., and Strauss, O. “A flexible targetless LiDAR–GNSS/INS–camera calibration method for UAV platforms”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 166 (2020), pp. 294–307.
- [PMP17] Peršić, J., Marković, I., and Petrović, I. “Extrinsic 6DoF calibration of 3D lidar and radar”. In: *2017 European Conference on Mobile Robots (ECMR)*. IEEE. 2017, pp. 1–6.
- [PMP19] Peršić, J., Marković, I., and Petrović, I. “Extrinsic 6dof calibration of a radar-lidar–camera system enhanced by radar cross section estimates evaluation”. In: *Robotics and Autonomous Systems* 114 (2019), pp. 217–230.
- [Qin+22] Qin, Z., Yu, H., Wang, C., Guo, Y., Peng, Y., and Xu, K. “Geometric transformer for fast and robust point cloud registration”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 11143–11152.

- [QY20] Quan, S. and Yang, J. “Compatibility-guided sampling consensus for 3-d point cloud registration”. In: *IEEE Transactions on Geoscience and Remote Sensing* 58.10 (2020), pp. 7380–7392.
- [Rey09] Reynolds, D. A. “Gaussian mixture models.” In: *Encyclopedia of biometrics* 741 (2009), pp. 659–663.
- [Rus19] Rusinkiewicz, S. “A symmetric objective function for ICP”. In: *ACM Transactions on Graphics (TOG)* 38.4 (2019), pp. 1–7.
- [RL01] Rusinkiewicz, S. and Levoy, M. “Efficient variants of the ICP algorithm”. In: *Proceedings third international conference on 3-D digital imaging and modeling*. IEEE. 2001, pp. 145–152.
- [RBB09] Rusu, R. B., Blodow, N., and Beetz, M. “Fast point feature histograms (FPFH) for 3D registration”. In: *2009 IEEE international conference on robotics and automation*. IEEE. 2009, pp. 3212–3217.
- [RC11] Rusu, R. B. and Cousins, S. “3d is here: Point cloud library (pcl)”. In: *2011 IEEE international conference on robotics and automation*. IEEE. 2011, pp. 1–4.
- [SSO20] Sahloul, H., Shirafuji, S., and Ota, J. “An accurate and efficient voting scheme for a maximally all-inlier 3D correspondence set”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.7 (2020), pp. 2287–2298.
- [STD14] Salti, S., Tombari, F., and Di Stefano, L. “SHOT: Unique signatures of histograms for surface and texture description”. In: *Computer Vision and Image Understanding* 125 (2014), pp. 251–264.
- [SDT09] Sandhu, R., Dambreville, S., and Tannenbaum, A. “Point set registration via particle filtering and stochastic dynamics”. In: *IEEE transactions on pattern analysis and machine intelligence* 32.8 (2009), pp. 1459–1473.
- [Sca11] Scaramuzza, D. “1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints”. In: *International journal of computer vision* 95 (2011), pp. 74–85.
- [Sch+19] Schöller, C., Schnettler, M., Krämmer, A., Hinz, G., Bakovic, M., Güzet, M., and Knoll, A. “Targetless rotational auto-calibration of radar and camera for intelligent transportation systems”. In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE. 2019, pp. 3934–3941.
- [Sch11] Scholz, D. *Deterministic global optimization: geometric branch-and-bound methods and their applications*. Vol. 63. Springer Science & Business Media, 2011.
- [SHT09] Segal, A., Haehnel, D., and Thrun, S. “Generalized-icp.” In: *Robotics: science and systems*. Vol. 2. 4. Seattle, WA. 2009, p. 435.
- [SMG23] Sidhartha, C., Manam, L., and Govindu, V. M. “Adaptive Annealing for Robust Geometric Estimation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 21929–21939.

- [SH06] Sim, K. and Hartley, R. “Removing outliers using the  $L_\infty$  norm”. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 1. IEEE. 2006, pp. 485–494.
- [SB11] Sipiran, I. and Bustos, B. “Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes”. In: *The Visual Computer* 27 (2011), pp. 963–976.
- [Sko80] Skolnik, M. I. “Introduction to radar systems”. In: *New York* (1980).
- [Str+17] Straub, J., Campbell, T., How, J. P., and Fisher, J. W. “Efficient global point cloud alignment using Bayesian nonparametric mixtures”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 2941–2950.
- [Sun+22a] Sun, C., Wei, Z., Huang, W., Liu, Q., and Wang, B. “Automatic Targetless Calibration for LiDAR and Camera Based on Instance Segmentation”. In: *IEEE Robotics and Automation Letters* (2022).
- [Sun21] Sun, L. “RANSIC: Fast and highly robust estimation for rotation search and point cloud registration using invariant compatibility”. In: *IEEE Robotics and Automation Letters* 7.1 (2021), pp. 143–150.
- [Sun+22b] Sun, Y., Li, J., Wang, Y., Xu, X., Yang, X., and Sun, Z. “ATOP: An attention-to-optimization approach for automatic LiDAR-camera calibration via cross-modal object matching”. In: *IEEE Transactions on Intelligent Vehicles* (2022).
- [Suw+18] Suwajanakorn, S., Snavely, N., Tompson, J. J., and Norouzi, M. “Discovery of latent 3d keypoints via end-to-end geometric reasoning”. In: *Advances in neural information processing systems* 31 (2018).
- [Svä+16] Svärm, L., Enqvist, O., Kahl, F., and Oskarsson, M. “City-scale localization for cameras with known vertical direction”. In: *IEEE transactions on pattern analysis and machine intelligence* 39.7 (2016), pp. 1455–1461.
- [TZN20] Tat-Jun, C., Zhipeng, C., and Neumann, F. “Robust Fitting in Computer Vision: Easy or Hard?” In: *International Journal of Computer Vision* 128.3 (2020), pp. 575–587.
- [TWS14] Theiler, P. W., Wegner, J. D., and Schindler, K. “Keypoint-based 4-points congruent sets—automated marker-less registration of laser scans”. In: *ISPRS journal of photogrammetry and remote sensing* 96 (2014), pp. 149–163.
- [TS+12] Theiler, P., Schindler, K., et al. “Automatic registration of terrestrial laser scanner point clouds using natural planar surfaces”. In: *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci* 3 (2012), pp. 173–178.
- [TN20] Tomazella, C. P. and Nagano, M. S. “A comprehensive review of Branch-and-Bound algorithms: Guidelines and directions for further research on the flow-shop scheduling problem”. In: *Expert Systems with Applications* 158 (2020), p. 113556.

- [TSD10] Tombari, F., Salti, S., and Di Stefano, L. “Unique signatures of histograms for local surface description”. In: *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part III 11*. Springer. 2010, pp. 356–369.
- [TSD13] Tombari, F., Salti, S., and Di Stefano, L. “Performance evaluation of 3D key-point detectors”. In: *International Journal of Computer Vision* 102.1 (2013), pp. 198–220.
- [TPH20] Tóth, T., Pusztai, Z., and Hajder, L. “Automatic LiDAR-camera calibration of extrinsic parameters using a spherical target”. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2020, pp. 8580–8586.
- [TK04] Tsing, Y. and Kanade, T. “A correlation-based approach to robust point set registration”. In: *European conference on computer vision*. Springer. 2004, pp. 558–569.
- [TAC19] Tzoumas, V., Antonante, P., and Carlone, L. “Outlier-robust spatial perception: Hardness, general-purpose algorithms, and guarantees”. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2019, pp. 5383–5390.
- [Wal+19] Walters, C., Mendez, O., Hadfield, S., and Bowden, R. “A robust extrinsic calibration framework for vehicles with unscaled sensors”. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2019, pp. 36–42.
- [Wan+21a] Wang, C., Liu, Y., Wang, Y., Li, X., and Wang, M. “Efficient and Outlier-Robust Simultaneous Pose and Correspondence Determination by Branch-and-Bound and Transformation Decomposition”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021), pp. 1–1. DOI: 10.1109/TPAMI.2021.3096842.
- [Wan+22] Wang, H., Liu, Y., Dong, Z., and Wang, W. “You only hypothesize once: Point cloud registration with rotation-equivariant descriptors”. In: *Proceedings of the 30th ACM International Conference on Multimedia*. 2022, pp. 1630–1641.
- [Wan+16] Wang, X., Xu, L., Sun, H., Xin, J., and Zheng, N. “On-Road Vehicle Detection and Tracking Using MMW Radar and Monovision Fusion”. In: *IEEE Transactions on Intelligent Transportation Systems* 17.7 (2016), pp. 2075–2084. DOI: 10.1109/TITS.2016.2533542.
- [Wan+21b] Wang, Y., Liu, Y., Li, X., Wang, C., Wang, M., and Song, Z. “Practical globally optimal consensus maximization by Branch-and-bound based on interval arithmetic”. In: *Pattern Recognition* 115 (2021), p. 107897.
- [WS19a] Wang, Y. and Solomon, J. M. “Deep closest point: Learning representations for point cloud registration”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 3523–3532.

- [WS19b] Wang, Y. and Solomon, J. M. “Prnet: Self-supervised learning for partial-to-partial registration”. In: *Advances in neural information processing systems* 32 (2019).
- [Wen+19] Wen, F., Ying, R., Gong, Z., and Liu, P. “Efficient algorithms for maximum consensus robust fitting”. In: *IEEE Transactions on Robotics* 36.1 (2019), pp. 92–106.
- [Wis+21] Wise, E., Peršić, J., Grebe, C., Petrović, I., and Kelly, J. “A Continuous-Time Approach for 3D Radar-to-Camera Extrinsic Calibration”. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2021, pp. 13164–13170.
- [Wu+20] Wu, J., Liu, M., Huang, Y., Jin, C., Wu, Y., and Yu, C. “SE (n)++: An efficient solution to multiple pose estimation problems”. In: *IEEE Transactions on Cybernetics* 52.5 (2020), pp. 3829–3840.
- [Xia+21] Xia, Y., Xu, Y., Li, S., Wang, R., Du, J., Cremers, D., and Stilla, U. “SOE-Net: A self-attention and orientation encoding network for point cloud based place recognition”. In: *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*. 2021, pp. 11348–11357.
- [Xie+18] Xie, S., Yang, D., Jiang, K., and Zhong, Y. “Pixels and 3-D points alignment method for the fusion of camera and LiDAR data”. In: *IEEE Transactions on Instrumentation and Measurement* 68.10 (2018), pp. 3661–3676.
- [Xu+22] Xu, W., Cai, Y., He, D., Lin, J., and Zhang, F. “Fast-lio2: Fast direct lidar-inertial odometry”. In: *IEEE Transactions on Robotics* 38.4 (2022), pp. 2053–2073.
- [Xu+19] Xu, Y., Boerner, R., Yao, W., Hoegner, L., and Stilla, U. “Pairwise coarse registration of point clouds in urban scenes using voxel-based 4-planes congruent sets”. In: *ISPRS journal of photogrammetry and remote sensing* 151 (2019), pp. 106–123.
- [Yan+22a] Yan, L., Wei, P., Xie, H., Dai, J., Wu, H., and Huang, M. “A New Outlier Removal Strategy Based on Reliability of Correspondence Graph for Fast Point Cloud Registration”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).
- [YC22] Yang, H. and Carlone, L. “Certifiably optimal outlier-robust geometric perception: Semidefinite relaxations and scalable global optimization”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.3 (2022), pp. 2816–2834.
- [YP23] Yang, H. and Pavone, M. “Object Pose Estimation with Statistical Guarantees: Conformal Keypoint Detection and Geometric Uncertainty Propagation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 8947–8958.
- [YSC20] Yang, H., Shi, J., and Carlone, L. “Teaser: Fast and certifiable point cloud registration”. In: *IEEE Transactions on Robotics* 37.2 (2020), pp. 314–333.



- [Yan+16] Yang, J., Li, H., Campbell, D., and Jia, Y. “Go-ICP: A Globally Optimal Solution to 3D ICP Point-Set Registration”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38.11 (2016), pp. 2241–2254. DOI: 10.1109/TPAMI.2015.2513405.
- [Yan+22b] Yang, J., Chen, J., Quan, S., Wang, W., and Zhang, Y. “Correspondence selection with loose–tight geometric voting for 3-D point cloud registration”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–14.
- [Yan+21] Yang, J., Huang, Z., Quan, S., Qi, Z., and Zhang, Y. “SAC-COT: Sample consensus by sampling compatibility triangles in graphs for 3-D point cloud registration”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021), pp. 1–15.
- [Yan+19] Yang, J., Xiao, Y., Cao, Z., and Yang, W. “Ranking 3D feature correspondences via consistency voting”. In: *Pattern Recognition Letters* 117 (2019), pp. 1–8.
- [Yan+23] Yang, J., Zhang, X., Fan, S., Ren, C., and Zhang, Y. “Mutual Voting for Ranking 3D Correspondences”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).
- [Ye+22] Ye, X., Shu, M., Li, H., Shi, Y., Li, Y., Wang, G., Tan, X., and Ding, E. “Rope3D: The Roadside Perception Dataset for Autonomous Driving and Monocular 3D Object Detection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (CVPR)*. 2022, pp. 21341–21350.
- [YL18] Yew, Z. J. and Lee, G. H. “3dfeat-net: Weakly supervised local 3d features for point cloud registration”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 607–623.
- [YL20] Yew, Z. J. and Lee, G. H. “Rpm-net: Robust point matching using learned features”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 11824–11833.
- [YL22] Yew, Z. J. and Lee, G. H. “Regtr: End-to-end point cloud correspondences with transformers”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 6677–6686.
- [YJC21] Yoon, B.-H., Jeong, H.-W., and Choi, K.-S. “Targetless Multiple Camera-LiDAR Extrinsic Calibration using Object Pose Estimation”. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2021, pp. 13377–13383.
- [Yu+21] Yu, H., Li, F., Saleh, M., Busam, B., and Ilic, S. “Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 23872–23884.
- [Yu+23] Yu, H., Qin, Z., Hou, J., Saleh, M., Li, D., Busam, B., and Ilic, S. “Rotation-invariant transformer for point cloud matching”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 5384–5393.

- [Yua+21] Yuan, C., Liu, X., Hong, X., and Zhang, F. “Pixel-level extrinsic self calibration of high resolution lidar and camera in targetless environments”. In: *IEEE Robotics and Automation Letters* 6.4 (2021), pp. 7517–7524.
- [Yua+20] Yuan, W., Eckart, B., Kim, K., Jampani, V., Fox, D., and Kautz, J. “Deepgmr: Learning latent gaussian mixture models for registration”. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V* 16. Springer. 2020, pp. 733–750.
- [Yuw+20] Yuwen, X., Chen, L., Yan, F., Zhang, H., Tang, J., Tian, B., and Ai, Y. “Improved vehicle LiDAR calibration with trajectory-based hand-eye method”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.1 (2020), pp. 215–224.
- [Zah+09] Zaharescu, A., Boyer, E., Varanasi, K., and Horaud, R. “Surface feature detection and description with applications to mesh matching”. In: *2009 IEEE conference on computer vision and pattern recognition*. IEEE. 2009, pp. 373–380.
- [Zen+17] Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., and Funkhouser, T. “3dmatch: Learning local geometric descriptors from rgb-d reconstructions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1802–1811.
- [ZS15] Zhang, J. and Singh, S. “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.
- [Zha+11] Zhang, J., Wang, F.-Y., Wang, K., Lin, W.-H., Xu, X., and Chen, C. “Data-driven intelligent transportation systems: A survey”. In: *IEEE Transactions on Intelligent Transportation Systems* 12.4 (2011), pp. 1624–1639.
- [ZYD21] Zhang, J., Yao, Y., and Deng, B. “Fast and robust iterative closest point”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.7 (2021), pp. 3450–3466.
- [ZZM15] Zhang, X., Zhang, A., and Meng, X. “Automatic fusion of hyperspectral images and laser scans using feature points”. In: *Journal of Sensors* 2015 (2015).
- [Zha+21] Zhang, X., Zhu, S., Guo, S., Li, J., and Liu, H. “Line-based automatic extrinsic calibration of lidar and camera”. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2021, pp. 9347–9353.
- [Zha+23a] Zhang, X., Peng, L., Xu, W., and Kneip, L. “Accelerating Globally Optimal Consensus Maximization in Geometric Vision”. In: *arXiv preprint arXiv:2304.05156* (2023).
- [Zha+23b] Zhang, X., Yang, J., Zhang, S., and Zhang, Y. “3D Registration with Maximal Cliques”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 17745–17754.

- [Zha+22] Zhao, M., Ma, L., Jia, X., Yan, D.-M., and Huang, T. “GraphReg: Dynamical Point Cloud Registration With Geometry-Aware Graph Signal Processing”. In: *IEEE Transactions on Image Processing* 31 (2022), pp. 7449–7464.
- [ZSO11] Zheng, Y., Sugimoto, S., and Okutomi, M. “Deterministically maximizing feasible subsystem for robust model fitting with unit norm constraint”. In: *CVPR 2011*. IEEE. 2011, pp. 1825–1832.
- [Zho09] Zhong, Y. “Intrinsic shape signatures: A shape descriptor for 3d object recognition”. In: *2009 IEEE 12th international conference on computer vision workshops, ICCV Workshops*. IEEE. 2009, pp. 689–696.
- [ZLK18] Zhou, L., Li, Z., and Kaess, M. “Automatic extrinsic calibration of a camera and a 3d lidar using line and plane correspondences”. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2018, pp. 5562–5569.
- [ZPK16] Zhou, Q.-Y., Park, J., and Koltun, V. “Fast global registration”. In: *European conference on computer vision*. Springer. 2016, pp. 766–782.
- [Zhu+22] Zhuang, G., Cagnetta, C., Bing, Z., Cao, H., Li, X., Huang, K., and Knoll, A. “A Biologically-Inspired Global Localization System for Mobile Robots Using LiDAR Sensor”. In: *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2022, pp. 984–990.
- [Zuñ+19] Zuñiga-Noël, D., Ruiz-Sarmiento, J.-R., Gomez-Ojeda, R., and Gonzalez-Jimenez, J. “Automatic multi-sensor extrinsic calibration for mobile robots”. In: *IEEE Robotics and Automation Letters* 4.3 (2019), pp. 2862–2869.