

Globally Optimal Consensus Maximization for Relative Pose Estimation With Known Gravity Direction

Yinlong Liu^{1b}, Guang Chen^{1b}, Rongqi Gu, and Alois Knoll^{1b}

Abstract—Relative pose estimation is a core task in robotic vision, and it is the basis of many high-level applications (e.g., visual odometry). In this letter, we focus on a quite common case in which the gravity direction is known in advance with the help of IMUs. Commonly, incorrect feature matches (a.k.a. outliers) are unavoidable, and they will impair the accuracy significantly. RANSAC is the de facto standard to suppress the outliers and obtain a robust solution. However, RANSAC is a non-deterministic algorithm, which means it produces a reasonable result only with a certain probability, and it cannot guarantee the global optimality to meet the safety demand in many life-critical applications. Therefore, we propose a globally optimal algorithm for relative pose estimation with known gravity direction. Specifically, the proposed method employs the branch-and-bound algorithm to solve a consensus maximization problem, and thus it is able to obtain the global solution with a provable guarantee. To verify the feasibility of our proposed method, both synthetic and real-data experiments are conducted. The experimental results support the global optimality of the proposed method and show that the method performs more robustly than existing methods.

Index Terms—SLAM, autonomous vehicle navigation, computer vision for automation.

I. INTRODUCTION

THE task of relative pose estimation is estimating the relative camera pose from matching correspondences of two frames, also known as essential matrix estimation [1]. It is one of core tasks in computer vision [2], thereby it is the basis of

Manuscript received February 17, 2021; accepted May 24, 2021. Date of publication June 7, 2021; date of current version June 18, 2021. This letter was recommended for publication by Associate Editor G. Gallego and Editor C. Cadena Lerma upon evaluation of the reviewers' comments. This work was supported by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3), and by the Key Technologies Development and Application of Piloted Autonomous Driving Trucks Project. (Corresponding author: Guang Chen.)

Yinlong Liu is with the Technical University of Munich, Munich 85748, Germany and also with Ningbo Hicren Biotechnology Co., Ltd., Ningbo 315336, China (e-mail: yinlong.liu@tum.de).

Guang Chen is with the Tongji University, Shanghai 200092, China and also with the Technical University of Munich, Munich 85748, Germany (e-mail: guang@in.tum.de).

Rongqi Gu is with the Autonomous Driving Group, Shanghai Westwell Information and Technology Co., Ltd., Shanghai 200050, China (e-mail: rongqi.gu@westwell-lab.com).

Alois Knoll is with the Technical University of Munich, Munich 85748, Germany (e-mail: knoll@in.tum.de).

Digital Object Identifier 10.1109/LRA.2021.3087080

many high-level applications (e.g., visual odometry, structure from motion and 3D reconstruction) [3]–[6].

In addition to cameras, modern applications, such as autonomous driving and robot navigation and localization, are usually equipped with many other sensors (e.g., GPS, inertial measurement units (IMUs)) [7], [8]. Therefore, we can obtain prior knowledge to help estimate the relative camera pose [9], [10].

In this letter, we focus on estimating relative pose with the prior known gravity direction [11]. The gravity direction can be usually provided by IMUs, and the accuracy of this direction is usually reliable even by low cost IMUs (typically error $< 0.5^\circ$) [12], [13]. Taking advantages of knowing gravity direction, the degrees-of-freedom of relative pose estimation problem reduce to three, which means three point correspondences instead of five [1] are sufficient to minimally estimate relative pose [11].

Unfortunately, the mismatches (a.k.a. outliers), which will impair the accuracy significantly [14], are unavoidable in real applications [9], [11]. To suppress the outliers, the defacto mechanism is *RANdom SAmple Consensus* (RANSAC) [15]. However, RANSAC is a non-deterministic algorithm, which means it cannot provide a correct solution with a provable guarantee [16]. More specifically, RANSAC produces a reasonable result only with a certain probability [14]. On the other hand, there are many safety-critical applications, which highly demand such algorithms that can return a extremely reliable solution in the presence of noise and outliers [17]. Obviously, RANSAC cannot meet this strict demand.

To provide a provably optimal solution for some safety-critical applications, we propose a globally-optimal solution, which applies a novel nested branch-and-bound (BnB) algorithm. The main contributions of this letter are as follows:

- In contrast to RANSAC, the proposed relative pose estimation method can obtain the globally-optimal solution with a provable guarantee, which means it can meet the strict demand in many safety-critical applications.
- A nested BnB algorithm with novel geometric bounds is proposed. More specifically, a new geometric formulation and its essential geometric relationship are explored.

II. RELATED WORK

We first review the outlier-free solutions for relative pose estimation with known gravity direction. In fact, the outlier-free

solutions are well studied [9], [11], [18], [19]. They focus on finding solutions to algebraic systems. Specifically, closed-form solutions for slightly different formulations are explored in [9] and [18], respectively. Furthermore, Sweeney *et al.* point out that solving for relative pose with known direction is a quadratic eigenvalue problem [11], and therefore, they propose a simple and extremely efficient algorithm. More recently, given prior knowledge of gravity direction, Ding *et al.* go deeper to explore minimal solutions to relative pose estimation problem with unknown focal length in [20]. Moreover, a non-minimal ($N \geq 4$) solution for relative pose estimation with gravity prior is explored in [21].

If the inputs are contaminated by outliers, outlier-free algorithms should be nested into RANSAC scheme, which is the de facto mechanism [9]. However, their solutions are sub-optimal due to the obvious heuristic nature of RANSAC [14]. To assure the global optimality of the optimal solution, Yang *et al.* [22] propose a globally optimal solution to essential matrix estimation. Specifically, they adopt the consensus set maximization as the objective and the branch-and-bound algorithm to systematically search for the global optimum. However, due to rather high dimensionality of the solution space (i.e., five degrees of freedom), Yang's BnB algorithm tends to be very slow [23]. Furthermore, Fredriksson *et al.* propose a globally optimal method for a more difficult case in which the correspondences are unknown [24]. In detail, they apply an efficient branch and bound technique in combination with bipartite matching to solve the correspondence problem. However, their method becomes intractable in the cases where outlier rate is considerably high and the number of input is considerably large. Besides, with the prior knowledge of full camera orientation, Fredriksson *et al.* explore globally-optimal methods for two-view translation estimation in [25] and [26]. Recently, under the plane-based Ackermann steering motion assumption, a globally optimal and correspondence-less solution is explored in [27].

In addition, the epipolar constraint can be linearized to enable the two-view relative pose to be estimated linearly [2]. Accordingly, many globally optimal algorithms for linear consensus maximization are proposed to solve the relative estimation problem. Li is one of pioneers to explore the globally-optimal solution for linearized relative pose estimation [28]. Furthermore, tree search method is proposed to find the global optimum in [29], and more recently tree search method is accelerated significantly in [30]. Nevertheless, because linearization will drop some constraints, the solution of linearized objective is not necessarily the same as the original solution.

It is worth mentioning that many recent studies focus on solving special pose estimation problems with the help of gravity direction. For example, homography-based minimal-case relative pose estimation with known gravity direction is thoroughly explored in [31]. Furthermore, minimal solutions for relative pose with a single affine correspondence and known vertical direction is studied in [32]. Besides, the prior knowledge of gravity direction is helpful for solving absolute pose estimation problem [8], [33] and PnL (*Perspective-n-Line*) problem [34], [35]. These gravity-known works also inspire our work.

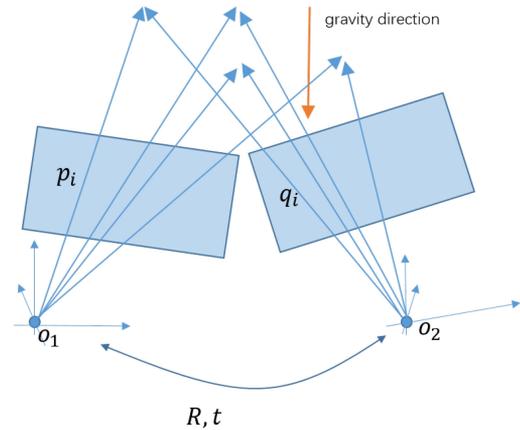


Fig. 1. The geometric of two-view relative pose estimation with the known gravity direction.

III. PROBLEM FORMULATION

Given N point correspondences $\{p_i, q_i\}_{i=1}^N$, which may be contaminated by outliers, the target is estimating the relative pose (i.e., rotation R and translation t). We start with the epipolar geometry [2] (see Fig. 1).

$$R\lambda_i^1 p_i + t = \lambda_i^2 q_i, i = 1 \dots N \quad (1)$$

where λ_i^1 and λ_i^2 are two different projective scales. Furthermore, Eq. (1) can be reformulated [11] as

$$t^T (q_i \times R p_i) = 0, i = 1 \dots N \quad (2)$$

where \times is cross product. It should mention that the scaling of t cannot be determined by Eq. (2) and we set $\|t\| = 1$.

In addition, given the unit gravity direction g_1 in left camera coordinate and g_2 in right camera coordinate, we have

$$g_2 = R g_1 \quad (3)$$

The solution of Eq. (3) is

$$R = R(\theta, g_2) \cdot R_{g_1}^{g_2} \quad (4)$$

where $R_{g_1}^{g_2}$ is the rotation that maps g_1 to g_2 with the minimum geodesic motion. $R(\theta, g_2)$ is the rotation that rotates θ about the axis g_2 and θ is the unknown-but-sought variable. Please refer to [36] for more details about the solution. Furthermore, according to Rodrigues' rotation formula [2],

$$R(\theta, g_2) = \exp(\theta [g_2]_{\times}) \quad (5)$$

$$= \mathbf{I} + \sin(\theta) [g_2]_{\times} + (1 - \cos(\theta)) [g_2]_{\times}^2 \quad (6)$$

where $[g_2]_{\times}$ is the cross-product matrix for g_2 .

For the i -th correspondence, with the help of gravity direction,

$$q_i \times R p_i = [q_i]_{\times} (\mathbf{I} + \sin(\theta) [g_2]_{\times} + (1 - \cos(\theta)) [g_2]_{\times}^2) R_{g_1}^{g_2} p_i \quad (7)$$

$$= a_i + \sin(\theta) b_i + \cos(\theta) c_i \quad (8)$$

where $c_i = -[q_i]_{\times} [g_2]_{\times}^2 R_{g_1}^{g_2} p_i$, $b_i = [q_i]_{\times} [g_2]_{\times} R_{g_1}^{g_2} p_i$ and $a_i = [q_i]_{\times} (\mathbf{I} + [g_2]_{\times}^2) R_{g_1}^{g_2} p_i$. Therefore, the epipolar constraint

becomes

$$\mathbf{t}^T (\mathbf{p}_i \times R\mathbf{q}_i) = 0 \quad (9)$$

$$\Rightarrow \mathbf{t}^T (\mathbf{a}_i + \sin(\theta) \mathbf{b}_i + \cos(\theta) \mathbf{c}_i) = 0 \quad (10)$$

In real application, it is almost impossible to estimate relative pose with clean and perfect point correspondences. The optimal solution should be obtained from the inlier observations, which should satisfy

$$|\mathbf{t}^{*T} (\mathbf{a}_i + \sin(\theta^*) \mathbf{b}_i + \cos(\theta^*) \mathbf{c}_i)| \leq \epsilon \quad (11)$$

where \mathbf{t}^* and θ^* are the optimal solution; ϵ is the inlier threshold. Therefore, a robust objective, which can suppress the outliers, should be formulated by utilizing the idea of inlier consensus maximization [14].

$$\max_{\mathbf{t}, \theta} \sum_{i=1}^N \mathbb{I} (|\mathbf{t}^T (\mathbf{a}_i + \sin(\theta) \mathbf{b}_i + \cos(\theta) \mathbf{c}_i)| \leq \epsilon) \quad (12)$$

where $\mathbb{I}(\cdot)$ is the indicator function which returns 1 if the condition \cdot is true and returns 0 if condition \cdot is not true.

To suppress the outliers, the robust objective is formulated. However, the objective is obviously non-smooth and non-concave which means some traditional optimizer (e.g., gradient descent) is infeasible and some (e.g., [37]) might be trapped in local optimum. Furthermore, many safety-critical applications need to obtain the globally optimal solution. To meet this demand, we apply the branch-and-bound algorithm, which is the most commonly used mechanism for solving NP-hard optimization problems [38], to seek the maximum value of the objective with provable guarantee.

IV. BRANCH-AND-BOUND

The BnB algorithm systematically explores the whole domain of the candidate solutions to find the optimal solution, thereby its solution is globally-optimal. More specifically, the BnB algorithm recursively branches the solution domain, and the sub-branches are checked against upper and lower estimated bounds on the optimal solution. If the branch cannot produce a better solution than the best one found so far by the algorithm, then it is discarded, and consequently, the solution domain is reduced. The iterative process of branching, bounding and cutting is terminated when the optimal solution is found.

A. Bounds Estimation

Obviously, the key of the BnB algorithm is how to estimate the upper and lower bounds in a given sub-branch. Accordingly, the solution domain should be parameterized properly. We note that the solution domain is $\mathbf{t} \in \mathbb{S}^2$ and $\theta \in [-\pi, \pi]$. In addition, if \mathbf{t} is the optimal solution then $-\mathbf{t}$ is also the optimal solution, therefore, we set the domain of \mathbf{t} a hemisphere \mathbb{S}^{2+} instead of a full sphere. \mathbb{S}^{2+} is defined as

$$\mathbb{S}^{2+} = \{\mathbf{t} | t_3 \geq 0\} \quad (13)$$

where $\mathbf{t} = [t_1, t_2, t_3]^T$ is a unit vector in \mathbb{R}^3 . To minimally parameterize the translation domain, we employ the exponential mapping method to map the hemisphere to a two dimensional

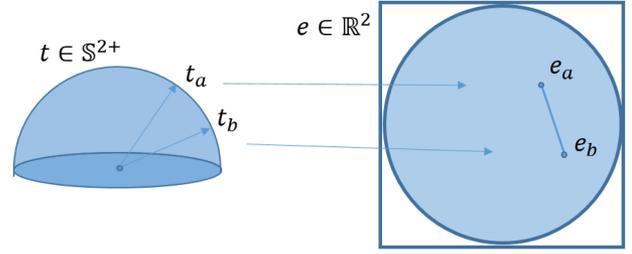


Fig. 2. The exponential mapping. Geometrically, the hemisphere is flattened into a disk in a plane, and we have $\angle(\mathbf{t}_a, \mathbf{t}_b) \leq \|\mathbf{e}_a - \mathbf{e}_b\|$.

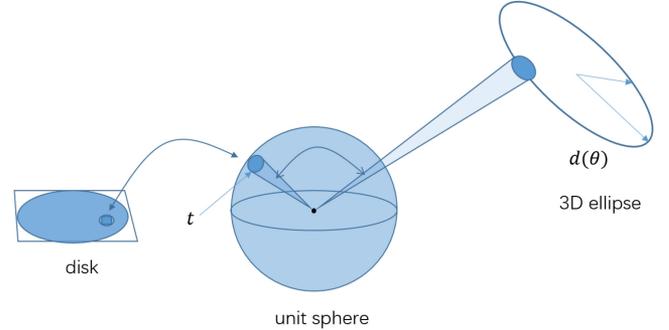


Fig. 3. The geometric of the upper bound. Geometrically, the solution domain of \mathbf{t} is a hemisphere, which can be mapped into a solid disk. Generally, $\mathbf{d}(\theta)$ is a 3D ellipse when $\theta \in [-\pi, \pi]$. Given the subdomain \mathbb{B}_t and \mathbb{B}_θ , \mathbf{t} should be in a relaxed umbrella-shaped area centered at \mathbf{t}_c and $\mathbf{d}(\theta)$ should be a curve which is in a relaxed ball centered at $\mathbf{d}(\theta_c)$.

disk [39]. More Specifically, given $\mathbf{t} \in \mathbb{S}^{2+}$, we can represent it by $\mathbf{e} \in \mathbb{R}^2$ (see Fig. 2).

$$\mathbf{t} = [\sin(\alpha) \cos(\beta), \sin(\alpha) \sin(\beta), \cos(\alpha)]^T \quad (14)$$

$$\mathbf{e} = \alpha [\cos(\beta), \sin(\beta)]^T \quad (15)$$

where $\alpha \in [0, \pi/2]$ and $\beta \in [-\pi, \pi]$. Consequently, we can apply the two dimensional disk to represent the solution domain of translation. Furthermore, for ease of manipulation, a circumscribed square of the solid disk is initialized as the translation solution domain in the BnB algorithm. Therefore, to find the optimal translation, the square-shaped branch will be subdivided into four sub-branches and we need to estimate the upper and lower bounds in these square-shaped sub-branches.

To derive the upper bound, we first introduce the following lemma (see [39]).

Lemma 1: Given $\mathbf{t}_a, \mathbf{t}_b \in \mathbb{S}^{2+}$, \mathbf{e}_a and \mathbf{e}_b are their corresponding points in the 2D disk. Then

$$\angle(\mathbf{t}_a, \mathbf{t}_b) \leq \|\mathbf{e}_a - \mathbf{e}_b\| \quad (16)$$

In the BnB algorithm, let $\mathbf{e} \in \mathbb{B}_t$, where \mathbb{B}_t is a translation branch. \mathbf{e}_c is the center of the square-shaped branch and δ is the half-side length. Then we have

$$\angle(\mathbf{t}, \mathbf{t}_c) \leq \|\mathbf{e} - \mathbf{e}_c\| \leq \sqrt{2}\delta \quad (17)$$

where \mathbf{t} and \mathbf{t}_c correspond to \mathbf{e} and \mathbf{e}_c , respectively.

Furthermore, we denote $\mathbf{d}_i(\theta) \triangleq \mathbf{a}_i + \sin(\theta) \mathbf{b}_i + \cos(\theta) \mathbf{c}_i$, and geometrically, $\mathbf{d}_i(\theta)$ is an ellipse (see Fig. 3). Let $\theta \in \mathbb{B}_\theta$ where \mathbb{B}_θ is an angle branch. θ_c is the center of the branch, and r_θ is the half length (radius) of the branch. Geometrically, when

$\theta \in \mathbb{B}_\theta$, $\mathbf{d}_i(\theta)$ is a curve of the full ellipse in 3D space. Inspired by *Lipschitz optimization* [40], [41] and *mean value inequality for vector-valued functions* [42], we have

$$\|\mathbf{d}_i(\theta) - \mathbf{d}_i(\theta_c)\| \leq r_\theta \cdot \max_{\theta \in \mathbb{B}_\theta} \|\mathbf{d}'_i(\theta)\| \quad (18)$$

where \mathbf{d}'_i is the first derivative of \mathbf{d}_i . More specifically,

$$\mathbf{d}'_i(\theta) = \cos(\theta)\mathbf{b}_i - \sin(\theta)\mathbf{c}_i \quad (19)$$

Then we define

$$\tau_i = r_\theta \cdot \overline{\|\mathbf{d}'_i(\theta)\|} \geq r_\theta \cdot \max_{\theta \in \mathbb{B}_\theta} \|\mathbf{d}'_i(\theta)\| \quad (20)$$

where $\overline{\|\mathbf{d}'_i(\theta)\|}$ is the upper bound of $\|\mathbf{d}'_i(\theta)\|$ and it can be calculated by *interval analysis* [43], [44]. Therefore,

$$\|\mathbf{d}_i(\theta) - \mathbf{d}_i(\theta_c)\| \leq \tau_i \quad (21)$$

Geometrically, given $\theta \in \mathbb{B}_\theta$, $\mathbf{d}_i(\theta)$ is a curve that is contained by a relaxed ball, whose center is at $\mathbf{d}_i(\theta_c)$ and radius is τ_i . (see Fig. 3)

$$\mathbf{d}_i(\theta) = \mathbf{d}_i(\theta_c) + \eta_i \mathbf{n} \quad (22)$$

where $0 \leq \eta_i \leq \tau_i$; \mathbf{n} is a unit direction. As a result,

$$\mathbf{t}^T \mathbf{d}_i(\theta_c) - \tau_i \leq \mathbf{t}^T \mathbf{d}_i(\theta) \leq \mathbf{t}^T \mathbf{d}_i(\theta_c) + \tau_i \quad (23)$$

We define $\xi_i = \angle(\mathbf{t}, \mathbf{d}_i(\theta_c))$. Since $\mathbf{t} \in \mathbb{S}^{2+}$, then

$$\|\mathbf{d}_i(\theta_c)\| \cos(\xi_i) - \tau_i \leq \mathbf{t}^T \mathbf{d}_i(\theta) \leq \|\mathbf{d}_i(\theta_c)\| \cos(\xi_i) + \tau_i \quad (24)$$

According to the triangle inequality in spherical geometry [45], we have

$$\angle(\mathbf{t}_c, \mathbf{d}_i(\theta_c)) - \angle(\mathbf{t}, \mathbf{t}_c) \leq \xi_i \leq \angle(\mathbf{t}_c, \mathbf{d}_i(\theta_c)) + \angle(\mathbf{t}, \mathbf{t}_c) \quad (25)$$

In addition, according to Eq. (17) we have

$$\angle(\mathbf{t}_c, \mathbf{d}_i(\theta_c)) - \sqrt{2}\delta \leq \xi_i \leq \angle(\mathbf{t}_c, \mathbf{d}_i(\theta_c)) + \sqrt{2}\delta \quad (26)$$

Consequently, we can sequentially derive the following bounds by *interval analysis*

$$[\xi_i] \stackrel{(26)}{\implies} [\cos(\xi_i)] \stackrel{(24)}{\implies} [\mathbf{t}^T \mathbf{d}_i(\theta)] \Rightarrow [|\mathbf{t}^T \mathbf{d}_i(\theta)|] \quad (27)$$

where $[\cdot]$ means calculating the upper and lower bound, and $[\cdot] \Rightarrow [\cdot]$ indicates that the bounds of the left-hand side leads to the bounds of the right-hand side. For simplicity, we define the lower bound of $|\mathbf{t}^T \mathbf{d}_i(\theta)|$ is $\Delta_i(\mathbb{B}_t, \mathbb{B}_\theta)$. Then given $e \in \mathbb{B}_t$ and $\theta \in \mathbb{B}_\theta$, the upper bound of the objective can be

$$U = \sum_{i=1}^N \mathbb{I}(\Delta_i(\mathbb{B}_t, \mathbb{B}_\theta) \leq \epsilon) \quad (28)$$

Proof: Since $\Delta_i(\mathbb{B}_t, \mathbb{B}_\theta) \leq |\mathbf{t}^T \mathbf{d}_i(\theta)|$, then

$$\mathbb{I}(\Delta_i(\mathbb{B}_t, \mathbb{B}_\theta) \leq \epsilon) \geq \mathbb{I}(|\mathbf{t}^T \mathbf{d}_i(\theta)| \leq \epsilon) \quad (29)$$

$$\Rightarrow \sum_{i=1}^N \mathbb{I}(\Delta_i(\mathbb{B}_t, \mathbb{B}_\theta) \leq \epsilon) \geq \max_{\mathbb{B}_t, \mathbb{B}_\theta} \sum_{i=1}^N \mathbb{I}(|\mathbf{t}^T \mathbf{d}_i(\theta)| \leq \epsilon) \quad (30)$$

Therefore, U is the upper bound. ■

For the lower bound, since the maximum value in the branch should not be less than the value at a specific point, it can be

Algorithm 1: Inner BnB: Finding the Optimal θ , Given a Translation Branch \mathbb{B}_t .

Input: Point Matching $\{\mathbf{p}_i, \mathbf{q}_i\}_{i=1}^N$, inlier threshold ϵ , gravity direction \mathbf{g}_1 and \mathbf{g}_2 and a translation branch \mathbb{B}_t .

Output: optimal θ^*

- 1 Initialize $\{\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i\}_{i=1}^N$;
- 2 Initialize a queue \mathbb{Q}_θ with a branch $[-\pi, \pi]$ and its initial upper and lower bounds are N and 0 ;
- 3 **while** \mathbb{Q}_θ is non empty **do**
- 4 Take out the best branch \mathbb{B}_θ which has the highest upper bound U_{in}^* ;
- 5 Branch \mathbb{B}_θ into sub-branches and estimate the upper and lower bounds;
- 6 Update the optimal θ^* with the center of the branch which has the highest lower bound L_{in}^* ;
- 7 Remove the branches whose upper bound is less than the highest lower bound;
- 8 **if** $L_{in}^* = U_{in}^*$ **then**
- 9 | Terminate
- 10 **end**
- 11 **end**

set as

$$L = \sum_{i=1}^N \mathbb{I}(\Delta_i(\mathbf{t}_c, \mathbf{d}_i(\theta_c)) \leq \epsilon) \quad (31)$$

where $\Delta_i(\mathbf{t}_c, \mathbf{d}_i(\theta_c))$ is the object value at a specific point, which means $\tau_i = 0$ and $\delta = 0$.

$$\xi_i \stackrel{\delta=0}{\implies} \cos(\xi_i) \stackrel{\tau_i=0}{\implies} \mathbf{t}_c^T \mathbf{d}_i(\theta_c) \Rightarrow |\mathbf{t}_c^T \mathbf{d}_i(\theta_c)| = \Delta_i(\mathbf{t}_c, \mathbf{d}_i(\theta_c)) \quad (32)$$

In addition, when the branch collapses to a specific point $\{\mathbf{t}_s, \theta_s\}$, then $\delta = 0$, $\xi_i = \angle(\mathbf{t}_s, \mathbf{d}_i(\theta_s))$ and $\tau_i = 0$. As a result, $|\mathbf{t}^T \mathbf{d}_i(\theta)| = |\mathbf{t}_s, \mathbf{d}_i(\theta_s)|$, then

$$L = U = \sum_{i=1}^N \mathbb{I}(|\mathbf{t}_s, \mathbf{d}_i(\theta_s)| \leq \epsilon) \quad (33)$$

In other words, the gap between upper and lower bound tends to be zero when the branch tends to a specific point. Thus the proposed upper and lower bounds are sufficient to be applied in BnB algorithm.

B. Nested BnB

To avoid heavily computational burden, we leverage on the nested BnB idea [46], which has better memory and computational efficiency [45]. Specifically, two BnB algorithms (one is for θ and the other is for \mathbf{t}) are executed in a nested manner to obtain the optimal solution.

Concretely, given $\mathbf{t} \in \mathbb{B}_t$ and $\theta \in [-\pi, \pi]$, we can use the BnB algorithm to find the optimal θ , which is named inner BnB. Accordingly, in the inner BnB, given a branch \mathbb{B}_θ , the upper bound is still

$$U_{in} = \sum_{i=1}^N \mathbb{I}(\Delta_i(\mathbb{B}_t, \mathbb{B}_\theta) \leq \epsilon) \quad (34)$$

Algorithm 2: Nested BnB: Finding the Optimal R, t .

Input: Point Matching $\{\mathbf{p}_i, \mathbf{q}_i\}_{i=1}^N$, inlier threshold ϵ , gravity direction \mathbf{g}_1 and \mathbf{g}_2 .
Output: optimal relative pose R^*, t^*

- 1 Initialize $\{\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i\}_{i=1}^N$;
- 2 Initialize a queue \mathbb{Q}_t with the solid square and its upper and lower bound are N and 0 ;
- 3 **while** \mathbb{Q}_t is not empty **do**
- 4 Take out the best branch \mathbb{B}_t which has the highest upper bound U_{out}^* ;
- 5 Branch \mathbb{B}_t into sub-branches;
- 6 Call Algorithm 1 to obtain the optimal θ^* in each sub-branch; /* \leftarrow Nested way \rightarrow */
- 7 Estimate the upper and lower bounds;
- 8 Update the optimal t^* with the center of the branch which has the highest lower bound L_{out}^* ;
- 9 Remove the branches whose upper bound is less than the highest lower bound;
- 10 **if** $L_{out}^* = U_{out}^*$ **then**
- 11 | Terminate
- 12 **end**
- 13 **end**

But the lower bound is slightly changed

$$L_{in} = \sum_{i=1}^N \mathbb{I}(\Delta_i(\mathbb{B}_t, \mathbf{d}_i(\theta_c)) \leq \epsilon) \quad (35)$$

where $\Delta_i(\mathbb{B}_t, \mathbf{d}_i(\theta_c))$ is the lower bound of $|\mathbf{t}^T \mathbf{d}_i(\theta_c)|$ at a specific angle θ_c , which means $\tau_i = 0$.

$$[\xi_i] \Rightarrow [\cos(\xi_i)] \xrightarrow{\tau_i=0} [\mathbf{t}^T \mathbf{d}_i(\theta_c)] \Rightarrow [|\mathbf{t}^T \mathbf{d}_i(\theta_c)|] \quad (36)$$

Furthermore, we can only consider branching the translation domain, which is called outer BnB. Accordingly, we denote the optimal θ^* which is obtained by the inner BnB in each branch. The upper and lower bound in the outer BnB are modified as

$$U_{out} = \sum_{i=1}^N \mathbb{I}(\Delta_i(\mathbb{B}_t, \mathbf{d}_i(\theta^*)) \leq \epsilon) \quad (37)$$

$$L_{out} = \sum_{i=1}^N \mathbb{I}(\Delta_i(\mathbf{t}_c, \mathbf{d}_i(\theta^*)) \leq \epsilon) \quad (38)$$

Note that L_{out} is the object value at a specific point (\mathbf{t}_c, θ^*) .

The outline of the inner BnB and the whole nested BnB algorithms are summarized in Algorithm 1 and Algorithm 2.

V. EXPERIMENTS

To verify the feasibility and the global optimality of our proposed method, we conduct experiments using both synthetic and real-world data. Besides, to demonstrate the performance, our proposed method and several state-of-the-art methods are compared.

A. Setup

All experiments are conducted in a computer with an AMD Ryzen 7 2700X CPU and 32 G RAM. In all the experiments, if not specified, the inlier threshold ϵ in the objective (Eq. (12)) is set to 0.001. All compared state-of-the-art algorithms are listed:

- RANSAC+3 pt: RANSAC framework with the state-of-the-art gravity-known three points algorithm [11]. The confidence $\rho = 0.999$ for the stopping criterion in all the experiments.
- RANSAC+5 pt: RANSAC framework with the famous five points algorithm [1]. The confidence $\rho = 0.999$ for the stopping criterion in all the experiments.
- 5d-BnB¹: the BnB method to solve the optimal essential matrix estimation problem without the prior knowledge of the gravity direction [22].
- A*²: A* tree search with Non-Adjacent Path Avoidance and Dimension-Insensitive Branch Pruning, which is named A*-NAPA-DIBP in original paper [30]. In addition, we set the maximum runtime to 60 seconds to avoid long time running.
- gBnB³: our proposed nested BnB method with the known gravity direction.

Note that 5d-BnB is written in C++, and other algorithms are run in MATLAB2020 A.

B. Controlled Experiments on Synthetic Data

In this section, to verify the global optimality of our proposed method, we conducted controlled experiments using synthetic data. First, we randomly generate 100 points in 3D world. Specifically, the 3D points are inside a cube whose side is 1 meter and center is about 2 meters away from the camera. A simulated camera with 1000 pixel focal length is randomly moved but still faces the 3D points. We then have 100 matching point pairs in 2D image plane. The virtual gravity is recorded in the two different coordinates. To simulate the mismatches (i.e., outliers), we replace the correct point matches by incorrect random point matches. The outlier rate is $\eta = N_{outlier}/N$ where $N_{outlier}$ is the number of outliers and N is the total number of inputs. Besides, to simulate the noise of point localization, we add the Gaussian noise to the data, and standard deviation σ is taken as the noise level. Moreover, to calculate the accuracy, we define the error as

$$e_r = \arccos(0.5(Tr(R'_{gt}R_{opt}) - 1)) \quad (39)$$

$$e_t = \angle(\mathbf{t}_{gt}, \mathbf{t}_{opt}) \quad (40)$$

where R_{gt} and \mathbf{t}_{gt} are motion ground truth; R_{opt} and \mathbf{t}_{opt} are estimated solution; $Tr(\cdot)$ is the trace function of a square matrix.

First, we test our proposed method in different outlier rates, $\eta = \{0, \dots, 0.5\}$. The noise level σ is set to 1. To observe the global optimality, the experiments are repeated 200 times for each experimental setting. We then calculate the success

¹<http://jlyang.org>

²<https://github.com/ZhipengCai/MaxConTreeSearch>

³<https://github.com/Liu-Yinlong/Globally-Optimal-Consensus-Maximization-for-Relative-Pose-Estimation-with-Known-Gravity-Direction>

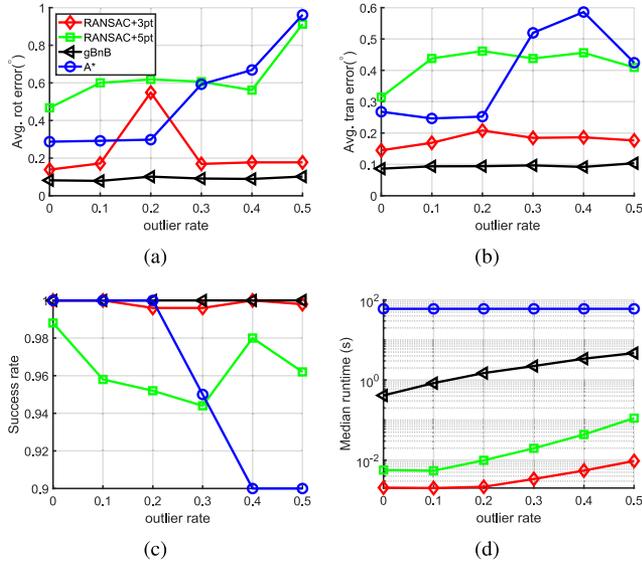


Fig. 4. Controlled experiments on synthetic data with different outlier rates.

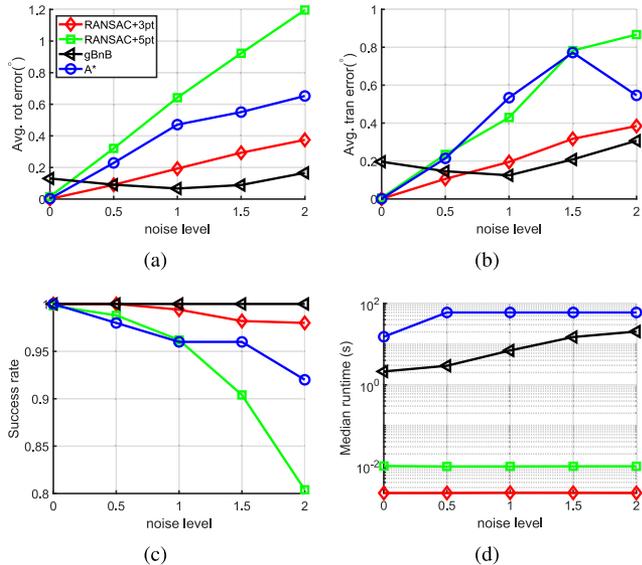


Fig. 5. Controlled experiments on synthetic data with different noise levels.

rate where a case that satisfies $err_{rot} \leq 2^\circ$ and $err_{tran} \leq 2^\circ$ is considered as a success case. In addition, the average error and median runtime are recorded. The results are showed in Fig. 4.

Furthermore, we test our proposed method in different noise levels, $\sigma = \{0, \dots, 2\}$. The outlier rate η is set to 0.2. 200 times are repeated in each experimental setting to observe the global optimality. The success rate, average error and median runtime are shown in Fig. 5.

In addition, we compare the efficiency between our proposed method and the 5d-BnB method, which does not rely on prior gravity [22]. In this experimental setting, no outliers and noise are added. We only test the methods in different number of inputs, $N = \{20, \dots, 140\}$. Since they are globally optimal methods, we only show the time duration in Table I, which is median runtime on 50 trials.

TABLE I
COMPARISON OF MEDIAN RUNTIME (SECONDS) ON SYNTHETIC DATA WITH DIFFERENT INPUT NUMBERS

point number	20	50	80	110	140
gBnB	0.287	0.351	0.392	0.462	0.534
5d-BnB	21.666	29.963	32.884	37.405	58.828

From all the results, we can draw the following points:

- Due to the random nature, RANSAC-based methods may return incorrect solutions while they run very fast. In contrast, our proposed BnB method can obtain the globally-optimal solution from outlier-contaminated inputs in this experimental setting, while it need more time than RANSAC-based methods.
- When outlier rate increases, tree search method consumes more runtime significantly. In addition, the tree search method cannot obtain all satisfactory solutions in our experiments. There are two reasons: (1) the optimal solution for linearized objective is not necessarily the same as that of original objective. (2) in large outlier rate, the tree search method reaches the time limitation (60 seconds) and terminates early, which mean it is not able to fully search all the solution candidates and just return the best-so-far solution. In contrast, our proposed method can obtain extremely robust solution and has the highest success rate using much less runtime.
- With the increase of noise level, the accuracy of all methods will decrease. However, when the noise level is large, our proposed method can obtain better accuracy than other methods, which reveals the robustness of our proposed method.
- With the help of gravity direction, our proposed BnB method is much faster than 5d-BnB method, which is one of the state-of-the-art globally-optimal methods. The reason is that with the help of gravity direction, the dimensionality of solution domain reduces from five to three, which leads to high efficiency of the BnB framework.

C. Robustness to IMU Noise

In this part, given a biased gravity direction, which simulates the measurement bias of IMUs, we conduct experiments to verify the robustness of our proposed method. First, we set $N = 100$, $\eta = 0.2$, $\sigma = 1$ and the biased angle is set from 0.1° to 0.5° . It is repeated 500 times under each experimental setting. In this case, we only compare our proposed method with RANSAC+3 pt, since other methods are not sensitive to gravity direction. The median runtime, average error and success rate are recorded. Besides, we also record the inlier number, which is a common metric in consensus maximization solutions [16], [30]. In general, the maximum objective for consensus maximization always occurs at the optimal solution, therefore, we record the consensus set number. The results are showed in Fig. 6.

From the results, we can find that when the gravity bias level increases, the accuracy of our proposed method decreases. It is reasonable because the given gravity direction is biased

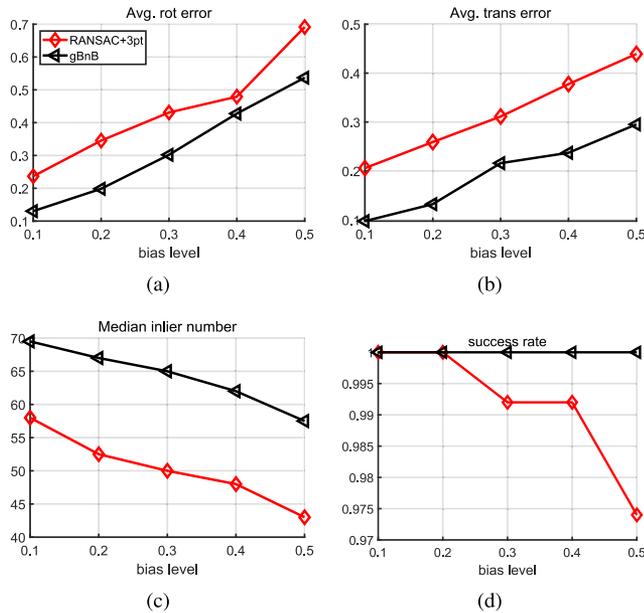
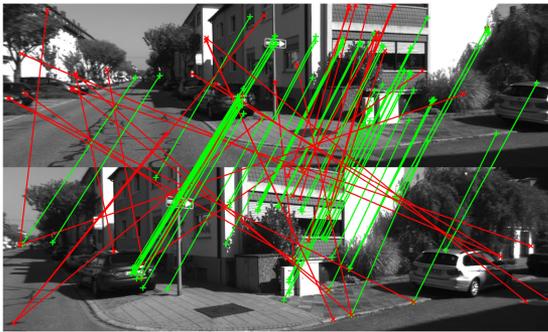
Fig. 6. Controlled experiments under different bias gravity levels ($^{\circ}$).

Fig. 7. Input correspondence (Frame 104-108) from KITTI dataset. Red lines denote the incorrect correspondences and green lines denote the correct inliers. The result is obtained by our proposed gBnB algorithm.

gradually. Nonetheless, our proposed method can still obtain satisfactory solution even with a biased gravity direction ($\leq 0.5^{\circ}$). It is worth noting that due to the biased gravity direction, RANSAC+3 pt method cannot obtain the maximum inlier number, and our proposed method is significantly more robust than the RANSAC+3 pt method.

D. Real-World Data Experiments

In this part, we verify the feasibility and practicality using real-world data. We select 5 image pairs (the first 5 crossroads) from the sequence 00 of the KITTI Odometry dataset [47], see Fig. 7. The image pairs are obtained under ground motion. This experimental setting is similar to the setting in [30]. We utilize MATLAB built-in functions *detectMSERFeatures*⁴ and *matchFeatures*⁵ to obtain the input correspondences. Note that to emphasize the outlier-robustness, we adjust the parameters

⁴<https://www.mathworks.com/help/vision/ref/detectmserfeatures.html>

⁵<https://www.mathworks.com/help/vision/ref/matchfeatures.html>

TABLE II
MAXIMUM ERROR AND AVERAGE RUNTIME(SECONDS) USING SELECTED IMAGES FROM KITTI DATASET IN 50 TIMES

Methods		RANSAC+3pt	RANSAC+5pt	A*	gBnB
Frame 104-108	e_r	1.438	1.248	3.592	0.685
	e_t	3.397	7.701	24.982	3.054
	T	1.749	21.162	60.028	23.706
Frame 198-201	e_r	0.389	0.672	2.295	0.346
	e_t	1.785	3.792	17.022	1.599
	T	1.703	20.448	60.076	20.078
Frame 417-420	e_r	0.218	0.268	0.806	0.176
	e_t	3.360	2.491	7.635	2.310
	T	1.761	20.333	18.106	5.566
Frame 579-582	e_r	0.276	0.375	1.452	0.242
	e_t	1.751	3.357	19.416	1.668
	T	1.757	20.392	60.166	50.020
Frame 738-742	e_r	0.653	0.817	0.997	0.597
	e_t	7.642	10.502	10.505	7.034
	T	1.745	20.215	54.302	12.632

to obtain more-than-usual mismatches. Besides, the iteration number of RANSAC-based methods is set to 10 000 since we have no prior knowledge of outlier rate in each scene.

To show the robustness, we repeat 50 times in each image pair and we record the maximum error. The results are listed in Table II. Note that 5d-BnB algorithm cannot terminate in 60 seconds in most of cases, we then do not list the results. From the results, we can find that RANSAC-based algorithms may return an unsatisfactory solution (error $> 10^{\circ}$). In addition, A* method may not return a correct solution due to time limitation and dropping non-linear constraints. In contrast, our proposed method usually obtain a satisfactory solution. It is worth mentioning that the proposed method needs more time than RANSAC-based method. However, the proposed method is more efficient than 5d-BnB and A* algorithm.

VI. CONCLUSION

In this letter, we focus on a special case that given the gravity direction, solving the relative pose from outlier-contaminated inputs. Since traditional RANSAC-based methods fail to guarantee the global optimality of the solution, we propose a novel nested BnB algorithm, which is able to obtain the globally-optimal solution. Even with a biased gravity direction, the proposed method can still obtain satisfactory solutions. We also verify the feasibility and practicality of our proposed method using both synthetic and real-world data. The experimental results reveal extreme robustness of our proposed method.

REFERENCES

- [1] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 756–770, Jun. 2004.
- [2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [3] R. Szeliski, *Computer Vision: Algorithms and Applications*. Berlin, Germany: Springer Science & Business Media, 2010.
- [4] T. Moons, L. Van Gool, and M. Vergauwen, *3D Reconstruction From Multiple Images: Principles*. Delft The Netherlands: Now Publishers Inc, 2009.
- [5] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: A versatile and accurate monocular slam system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.

- [6] G. Chen, H. Cao, J. Conradt, H. Tang, F. Rohrbein, and A. Knoll, "Event-based neuromorphic vision for autonomous driving: A paradigm shift for bio-inspired visual sensing and perception," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 34–49, Jul. 2020.
- [7] O. Saurer, P. Vasseur, R. Bouteau, C. Démonceaux, M. Pollefeys, and F. Fraundorfer, "Homography based egomotion estimation with a common direction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 327–341, Feb. 2017.
- [8] L. Svärm, O. Enqvist, F. Kahl, and M. Oskarsson, "City-scale localization for cameras with known vertical direction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1455–1461, Jul. 2017.
- [9] O. Naroditsky, X. S. Zhou, J. Gallier, S. I. Roumeliotis, and K. Daniilidis, "Two efficient solutions for visual odometry using directional correspondence," *IEEE Trans. Pattern Anal. Mach. Int.*, vol. 34, no. 4, pp. 818–824, Apr. 2012.
- [10] M. Kalantari, A. Hashemi, F. Jung, and J.-P. Guédon, "A new solution to the relative orientation problem using only 3 points and the vertical direction," *J. Math. Imag. Vision*, vol. 39, no. 3, pp. 259–268, 2011.
- [11] C. Sweeney, J. Flynn, and M. Turk, "Solving for relative pose with a partially known rotation is a quadratic eigenvalue problem," in *Proc. 2nd Int. Conf. 3D Vision*, vol. 1. IEEE, 2014, pp. 483–490.
- [12] Z. Kukulova, M. Bujnak, and T. Pajdla, "Closed-form solutions to minimal absolute pose problems with known vertical direction," in *Proc. Asian Conf. Comput. Vision*, Springer, 2010, pp. 216–229.
- [13] N. Horanyi and Z. Kato, "Generalized pose estimation from line correspondences with known vertical direction," in *Proc. Int. Conf. 3D Vision, (3DV)*. IEEE, 2017, pp. 244–253.
- [14] T.-J. Chin and D. Suter, "The maximum consensus problem: Recent algorithmic advances," *Synth. Lectures Comput. Vision*, vol. 7, no. 2, pp. 1–194, 2017.
- [15] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [16] H. M. Le, T.-J. Chin, A. Eriksson, T.-T. Do, and D. Suter, "Deterministic approximate methods for maximum consensus robust fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 842–857, Mar. 2021.
- [17] V. Tzoumas, P. Antonante, and L. Carlone, "Outlier-robust spatial perception: Hardness, general-purpose algorithms, and guarantees," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 5383–5390.
- [18] F. Fraundorfer, P. Tanskanen, and M. Pollefeys, "A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles," in *Proc. Eur. Conf. Comput. Vision*, Springer, 2010, pp. 269–282.
- [19] Y. Ding, J. Yang, and H. Kong, "An efficient solution to the relative pose estimation with a common direction," in *Proc. IEEE Int. Conf. Robot. Automat.*. IEEE, 2020, pp. 11 053–11 059.
- [20] Y. Ding, J. Yang, J. Ponce, and H. Kong, "Minimal solutions to relative pose estimation from two views sharing a common direction with unknown focal length," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, 2020.
- [21] Y. Ding, D. Barath, J. Yang, H. Kong, and Z. Kukulova, "Globally optimal relative pose estimation with gravity prior," Dec. 2020, *arXiv:2012.00458*.
- [22] J. Yang, H. Li, and Y. Jia, "Optimal essential matrix estimation via inlier-set maximization," in *Eu. Conf. Comput. Vision*, Springer, 2014, pp. 111–126.
- [23] K. Joo, H. Li, T.-H. Oh, Y. Bok, and I. S. Kweon, "Globally optimal relative pose estimation for camera on a selfie stick," in *Proc. IEEE Int. Conf. Robot. Automat.*. IEEE, 2020, pp. 4983–4989.
- [24] J. Fredriksson, V. Larsson, C. Olsson, and F. Kahl, "Optimal relative pose with unknown correspondences," in *Proc. IEEE Conf. Comput. Vision, Pattern Recognit.*, 2016, pp. 1728–1736.
- [25] J. Fredriksson, O. Enqvist, and F. Kahl, "Fast and reliable two-view translation estimation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2014.
- [26] J. Fredriksson, V. Larsson, and C. Olsson, "Practical robust two-view translation estimation," in *Proc. IEEE Conf. Comput. Vision, Pattern Recognit.*, 2015, pp. 2684–2690.
- [27] L. Gao, J. Su, J. Cui, X. Zeng, X. Peng, and L. Kneip, "Efficient globally-optimal correspondence-less visual odometry for planar ground vehicles," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 2696–2702.
- [28] H. Li, "Consensus set maximization with guaranteed global optimality for robust geometry estimation," in *Proc. IEEE 12th Int. Conf. Comput. Vision*, IEEE, 2009, pp. 1074–1080.
- [29] T.-J. Chin, P. Purkait, A. Eriksson, and D. Suter, "Efficient globally optimal consensus maximisation with tree search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 758–772, Apr. 2017, doi: [10.1109/TPAMI.2016.2631531](https://doi.org/10.1109/TPAMI.2016.2631531).
- [30] Z. Cai, T.-J. Chin, and V. Koltun, "Consensus maximization tree search revisited," in *Proc. IEEE Int. Conf. Comput. Vision*, 2019, pp. 1637–1645.
- [31] Y. Ding, J. Yang, J. Ponce, and H. Kong, "Homography-based minimal-case relative pose estimation with known gravity direction," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020, doi: [10.1109/TPAMI.2020.3005373](https://doi.org/10.1109/TPAMI.2020.3005373).
- [32] B. Guan, J. Zhao, Z. Li, F. Sun, and F. Fraundorfer, "Minimal solutions for relative pose with a single affine correspondence," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, 2020.
- [33] C. Albl, Z. Kukulova, and T. Pajdla, "Rolling shutter absolute pose problem with known vertical direction," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016.
- [34] L. Lecrosnier, R. Bouteau, P. Vasseur, X. Savatier, and F. Fraundorfer, "Camera pose estimation based on pnl with a known vertical direction," *IEEE Robot. Automat. Lett.*, vol. 4, no. 4, pp. 3852–3859, Oct. 2019.
- [35] H. Abdellali, R. Frohlich, and Z. Kato, "Robust absolute and relative pose estimation of a central camera system from 2d-3 d line correspondences," in *Proc. IEEE/CVF Int. Conf. Comput. Vision Workshops*, 2019.
- [36] Á. P. Bustos and T.-J. Chin, "Guaranteed outlier removal for point cloud registration with correspondences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2868–2882, Dec. 2018.
- [37] Z. Cai, T.-J. Chin, H. Le, and D. Suter, "Deterministic consensus maximization with biconvex programming," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 685–700.
- [38] J. Clausen, "Branch and bound algorithms-principles and examples," Department of Computer Science. København, Denmark: Univ. of Copenhagen, pp. 1–30, 1999.
- [39] Y. Liu, G. Chen, and A. Knoll, "Globally optimal vertical direction estimation in atlanta world," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020, doi: [10.1109/TPAMI.2020.3027047](https://doi.org/10.1109/TPAMI.2020.3027047).
- [40] H. Li and R. Hartley, "The 3d-3 d registration problem revisited," in *Proc. IEEE 11th Int. Conf. Comput. Vision*, IEEE, 2007, pp. 1–8.
- [41] P. Hansen and B. Jaumard, "Lipschitz Optimization," in *Handbook of Global Optimization*. New York: Springer, 1995, pp. 407–493.
- [42] *Mean Value Theorem*. [Online]. Available: https://en.wikipedia.org/wiki/Mean_value_theorem
- [43] E. Hansen and G. W. Walster, *Global Optimization Using Interval Analysis: Revised and Expanded*. Boca Raton, FL, USA: CRC Press, 2003, vol. 264.
- [44] J.-C. Bazin *et al.*, "Globally optimal line clustering and vanishing point estimation in manhattan world," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*. IEEE, 2012, pp. 638–645.
- [45] D. Campbell, L. Petersson, L. Kneip, and H. Li, "Globally-optimal inlier set maximisation for camera pose and correspondence estimation," *IEEE Trans. Pattern Anal. Mach. Int.*, vol. 42, no. 2, pp. 328–342, Feb. 2020.
- [46] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-Icp: A globally optimal solution to 3 d Icp point-set registration," *IEEE Trans. Pattern Anal. Mach. Int.*, vol. 38, no. 11, pp. 2241–2254, Nov. 2016, doi: [10.1109/TPAMI.2015.2513405](https://doi.org/10.1109/TPAMI.2015.2513405).
- [47] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*. IEEE, 2012, pp. 3354–3361.