Quasi-globally Optimal and Efficient Vanishing Point Estimation in Manhattan World

Haoang Li[1] Ji Zhao[2] Jean-Charles Bazin[3] Wen Chen[1] Zhe Liu[1] Yun-Hui Liu[1]
[1]The Chinese University of Hong Kong, China [2]TuSimple, China [3]KAIST, South Korea
{hali, wenchen, zliu, yhliu}@mae.cuhk.edu.hk zhaoji84@gmail.com bazinjc@kaist.ac.kr

Abstract

The image lines projected from parallel 3D lines intersect at a common point called the vanishing point (VP). Manhattan world holds for the scenes with three orthogonal VPs. In Manhattan world, given several lines in a calibrated image, we aim at clustering them by three unknown-but-sought VPs. The VP estimation can be reformulated as computing the rotation between the Manhattan frame and the camera frame. To compute this rotation, state-of-the-art methods are based on either data sampling or parameter search, and they fail to guarantee the accuracy and efficiency simultaneously. In contrast, we propose to hybridize these two strategies. We first compute two degrees of freedom (DOF) of the above rotation by two sampled image lines, and then search for the optimal third DOF based on the branch-and-bound. Our sampling accelerates our search by reducing the search space and simplifying the bound computation. Our search is not sensitive to noise and achieves quasi-global optimality in terms of maximizing the number of inliers. Experiments on synthetic and real-world images showed that our method outperforms state-of-the-art approaches in terms of accuracy and/or efficiency.

1. Introduction

Vanishing point (VP) is the intersection of a set of image lines projected from parallel 3D lines. It has been successfully applied to various fields such as structure from motion [15, 23], scene understanding [14] and SLAM [21, 22]. Structured environments (typically man-made scenes) exhibit particular regularity like parallelism and orthogonality. Manhattan world [11] holds for the scenes with three mutually orthogonal dominant directions that correspond to three orthogonal VPs (shown in Fig. 1). In Manhattan world, given several lines in a calibrated image, we aim at clustering them by three unknown-but-sought VPs.

The Manhattan frame (MF) [30] is widely used to model the structure of Manhattan world. Three axes of MF correspond to three dominant directions of Manhattan world.

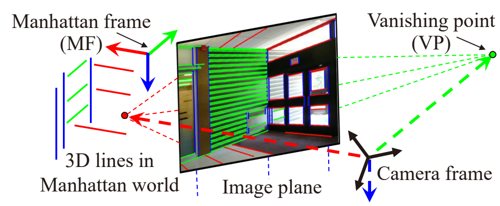

Figure 1. We cluster a set of image lines into three groups (shown in different colors) by three unknown-but-sought orthogonal VPs. The VP estimation can be reformulated as computing the rotation between the MF and the camera frame.

Fig. 1 shows that the camera center and VPs define three orthogonal 3D directions aligned to the MF axes. Based on this constraint, the VP estimation can be reformulated as computing the rotation between the MF and the camera frame (hereinafter called the "MF rotation"). To compute this rotation, state-of-the-art methods are based on either data sampling [4, 25, 35] or parameter search [5, 6, 24]. They hypothesize MF rotations in different ways and retrieve the optimal one fitting most image lines. The sampling-based methods (typically using RANSAC [13]) hypothesize finite MF rotations by the sampled image lines. Due to the sampling uncertainty, they fail to guarantee the global optimality in terms of maximizing the number of inliers. The search-based methods (typically using branch-and-bound (BnB) [20]) directly face infinite MF rotations hypothesized over the rotation space, and continuously narrow down the search scope. While they achieve global optimality in terms of maximizing the number of inliers, their efficiency is limited by high-dimensional search space and complex computation of the bounds of cost functions.

In contrast, we propose to hybridize the sampling and search strategies to compute three degrees of freedom (DOF) of the MF rotation. We first efficiently estimate two DOF by two sampled image lines, and then employ BnB to search for the optimal third DOF. The key advantage of our method is that it achieves both high efficiency and high accuracy. For efficiency, we exploit sampling to accelerate our search by reducing the search space. Moreover, our rotation parametrization contributes to efficiently computing tight bounds of our cost function, which further speeds up our search. Therefore, our method is more efficient than the pure search-based approaches [5, 6]. For accuracy, we search for the optimal third DOF that maximizes the number of inliers. Thanks to our search, our estimated MF rotation can be treated as the "quasi-globally" optimal solution in terms of number of inliers. While the global optimality may not be achieved due to our sampling, our method is less sensitive to noise and retrieves more inliers than the pure sampling-based approaches [4, 25, 35].

Overall, we propose a quasi-globally optimal and efficient VP estimation method by hybridizing sampling and search strategies. Our main contributions are:

- We leverage two sampled image lines to efficiently compute two DOF of the MF rotation. Our sampling accelerates our search by reducing the search space and simplifying the bound computation.
- We exploit BnB to search for the optimal third DOF of the MF rotation by fixing the other two DOF, achieving the quasi-global optimality. Our search is not sensitive to noise and obtains a large number of inliers.
- For the cases that the pure search-based methods fail to handle, our approach provides correct VPs and also is more accurate than the pure sampling-based ones.

Experiments showed that our method outperforms state-of-the-art approaches in terms of accuracy and/or efficiency.

2. Related Work

Existing VP estimation methods can be classified into four main categories with respect to the used algorithms, i.e. Hough transform [2, 28], expectation-maximization [1, 12], data sampling [4, 25, 31, 35] and parameter search [3, 5, 6].

The Hough transform-based methods [2,28] compute the intersections of all pairs of image lines and generate a histogram of these intersections. The bins with large numbers of entries correspond to VPs. However, they often lead to multiple and/or false detections, and also neglect the orthogonality constraint of VPs. In addition, the expectation-maximization-based methods [1, 12] cluster image lines and estimate VPs alternately. They assign each image line with a label indicating which cluster it belongs to. They use the lines with the same label to compute VPs, which in turn updates labels. However, they are sensitive to the initial solution and prone to converging to a local optimum.

The data sampling-based methods [4, 25, 31, 35] exploit RANSAC [13] and its variants [32, 36]. State-of-the-art ones [4, 25, 35] first sample three image lines several times to hypothesize finite VP triplets or MF rotations (earlier works like [29] hypothesize VPs individually and have lower accuracy and efficiency). Then they test each hypothesis by counting the number of image lines fitting this hypothesis. The fitness represents that the image line passes a VP or its associated projection plane normal is orthogonal to a MF axis up to a threshold. After that, they retrieve

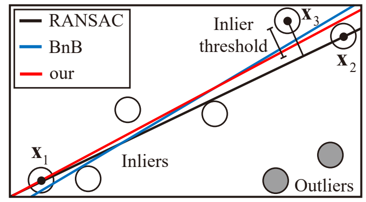

Figure 2. Illustration of different methods to obtain the optimal line fitting most inliers, given the input points $\{\mathbf{x}_i\}$ corrupted by noise and outliers. The RANSAC line must pass through two points (here $\mathbf{x}_1$ and $\mathbf{x}_2$), and treats $\mathbf{x}_3$ as an outlier since its distance is higher than the threshold, while the BnB line and our line both treat $\mathbf{x}_3$ as an inlier.

the hypothesis fitting most inliers. In addition, Tardif [31] used numerous hypotheses to define the image line descriptors and clustered lines by J-Linkage [32], a variant of RANSAC. However, it fails to enforce the orthogonality of VPs when clustering image lines. Note that the above sampling-based methods cannot guarantee the global optimality in terms of maximizing the number of inliers due to the sampling uncertainty.

The parameter search-based methods typically employ BnB [5, 6]. They directly face infinite hypothesized MF rotations over the rotation space parametrized by Euler angles or axis–angle representation. They search for the optimal rotation fitting most image lines by continuously narrowing down the search scope. While they guarantee the global optimality, their efficiency is unsatisfactory (more than 5 seconds per image in general). Joo et al. [18] recently proposed a novel strategy to significantly improve the efficiency of BnB, but it is not well applicable to image lines (it is inherently suitable for 3D plane normals). In addition, Bazin et al. [3] proposed to sample numerous MF rotations over the rotation space (i.e. the quasi-exhaustive search) and select the one maximizing the number of inliers. While it is appropriate for smooth videos, it can be computationally expensive, especially for single images or when no prior information regarding the camera orientation is available.

Overall, state-of-the-art VP estimation methods based on the sampling or search cannot guarantee the accuracy and efficiency simultaneously. In contrast, we propose to hybridize these two strategies, achieving high accuracy and high efficiency. Moreover, our method enforces the VP orthogonality thanks to the orthogonality of the MF rotation.

3. Algorithm Overview

We begin with taking the 2D line fitting for example to show the idea and advantage of our method. As shown in Fig. 2, given a set of points, we aim at obtaining the optimal line that fits most inliers (the direct least-squares fitting is not suitable due to the presence of outliers). We express a line by a point lying on it (i.e. the line position) and the line direction. RANSAC [13] hypothesizes a set of lines by sampling two points several times. The line defined by the points $\mathbf{x}_1$ and $\mathbf{x}_2$ is its optimal hypothesis, but fails to fit the inlier point $\mathbf{x}_3$. In addition, BnB [20] searches over the parameter spaces with respect to both line position and line direction (which correspond to infinite hypothesized lines). While BnB obtains the globally optimal line fitting all the inliers, its efficiency is unsatisfactory due to its high-dimensional search space. In contrast, we propose to first sample only one point to fix the line position and then search over the line direction space. We obtain the optimal line that passes through the point $\mathbf{x}_1$ and also fits all the inliers. In a sense, our strategy is "in between", i.e. we hybridize the data sampling and parameter search strategies. For efficiency, we leverage sampling to accelerate our search by reducing the full search space to the line direction space. For accuracy, we search for the optimal line direction that maximizes the number of inliers.

To compute the MF rotation, we propose a hybrid method to estimate its three DOF. Specifically, given a set of image lines, we first estimate two DOF by two sampled image lines in Section 4. It is similar to the above estimation of the line position. Then we search for the optimal third DOF maximizing the number of inlier image lines in Section 5. It is similar to the above search of the line direction.

4. Computing Two DOF by Sampling

In this section, we propose a novel method to compute two DOF of the MF rotation by two sampled image lines, and parametrize the MF rotation by a single parameter.

4.1. Two-line MF Rotation Parametrization

We assume the intrinsic matrix of the camera is known from calibration [17]. As shown in Fig. 3, to compute two DOF of the MF rotation, we leverage two sampled image lines $\{\mathbf{l}_1, \mathbf{l}_2\}$ whose associated 3D lines $\{\mathbf{L}_1, \mathbf{L}_2\}$ are aligned to two MF axes, i.e. $\mathbf{l}_1$ and $\mathbf{l}_2$ correspond to different VPs. We normalize the image by the intrinsic matrix [35] to compute 1) the normal $\mathbf{n}_1$ of the projection plane $\boldsymbol{\pi}_1$ by $\mathbf{l}_1$, and 2) the bases $\{\mathbf{s}_2, \mathbf{e}_2\}$ of the projection plane $\boldsymbol{\pi}_2$ by the endpoints of $\mathbf{l}_2$. We parametrize the unknown 3D line directions $\{\mathbf{d}_1, \mathbf{d}_2\}$ of $\{\mathbf{L}_1, \mathbf{L}_2\}$ as follows.

We first consider the 3D line direction $\mathbf{d}_1$. We define the direction $\mathbf{w}$ orthogonal to the projection plane normal $\mathbf{n}_1$, i.e. $\mathbf{n}_1^\top \mathbf{w} = 0$, and set $\mathbf{w}$ as any unit basis of the null space of $\mathbf{n}_1$ ($\mathbf{w}$ is not unique). Then we rotate $\mathbf{w}$ around the known rotation axis $\mathbf{n}_1$ by the unknown-but-sought angle $\theta \in [0, \pi]$ to align $\mathbf{w}$ to the direction $\mathbf{d}_1$ as

$$\mathbf{d}_1 = \breve{\mathbf{R}}(\mathbf{n}_1, \theta)\mathbf{w}, \tag{1}$$

where $\breve{\mathbf{R}}(\cdot,\cdot)$ denotes the axis-angle representation [17]. Based on Eq. (1), we parametrize $\mathbf{d}_1$ by θ. Each element of $\mathbf{d}_1$ is expressed by $d_{1,i}(\theta) = \boldsymbol{\omega}_{1,i}^\top \boldsymbol{\alpha}$ $(i = 1, 2, 3)$ where $\boldsymbol{\omega}_{1,i}$ is a known 2D vector and $\boldsymbol{\alpha} = [\sin(\theta), \cos(\theta)]^\top$. The norm of $\mathbf{d}_1$ is 1 since it is rotated from the unit vector $\mathbf{w}$.

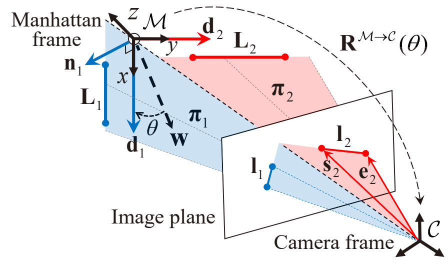

Figure 3. The image lines $\{\mathbf{l}_1, \mathbf{l}_2\}$ are associated with the 3D lines $\{\mathbf{L}_1, \mathbf{L}_2\}$ aligned to two MF axes. We use $\{\mathbf{l}_1, \mathbf{l}_2\}$ to parametrize the MF rotation $\mathbf{R}^{\mathcal{M}\to\mathcal{C}}$ by the unknown-but-sought angle θ.

Then we parametrize the 3D line direction $\mathbf{d}_2$ by enforcing two constraints. As shown in Fig. 3, first, $\mathbf{d}_2$ is parallel to the plane $\boldsymbol{\pi}_2$. We thus express $\mathbf{d}_2$ as a linear combination of the known bases $\{\mathbf{s}_2, \mathbf{e}_2\}$ of $\boldsymbol{\pi}_2$, i.e. $\mathbf{d}_2 = \mathbf{s}_2 + \lambda \cdot \mathbf{e}_2$ where λ is the unknown combination coefficient. Second, the 3D line directions $\mathbf{d}_1$ and $\mathbf{d}_2$ are mutually orthogonal, i.e. $\mathbf{d}_1^\top \mathbf{d}_2 = 0$. We combine these two constraints to express the coefficient λ by $\lambda = -(\mathbf{d}_1^\top \mathbf{s}_2)/(\mathbf{d}_1^\top \mathbf{e}_2)$. Then we substitute λ back into the first constraint and obtain $\mathbf{d}_2$ as

$$\mathbf{d}_2 = \mathbf{s}_2 - \frac{\mathbf{d}_1^\top \mathbf{s}_2}{\mathbf{d}_1^\top \mathbf{e}_2}\mathbf{e}_2 \propto \mathbf{d}_1^\top \mathbf{e}_2 \mathbf{s}_2 - \mathbf{d}_1^\top \mathbf{s}_2 \mathbf{e}_2, \tag{2}$$

where "$\propto$" denotes the equality regardless of scale. By substituting Eq. (1) into Eq. (2), we can parametrize $\mathbf{d}_2$ by θ. Each element of $\mathbf{d}_2$ is expressed by $d_{2,i}(\theta) = \boldsymbol{\omega}_{2,i}^\top \boldsymbol{\alpha}$ $(i = 1, 2, 3)$ where $\boldsymbol{\omega}_{2,i}$ is a known 2D vector. Note that the elements of $\mathbf{d}_1$ and $\mathbf{d}_2$ are both composed of $\boldsymbol{\alpha} = [\sin(\theta), \cos(\theta)]^\top$.

Let $\mathbf{d}_3$ denote the 3D direction orthogonal to both $\mathbf{d}_1$ and $\mathbf{d}_2$, and we compute it by $\mathbf{d}_3 = \mathbf{d}_1 \times \mathbf{d}_2$. Based on Eqs. (1) and (2), $\mathbf{d}_3$ is also parametrized by θ. Each element of $\mathbf{d}_3$ is expressed by $d_{3,i}(\theta) = \boldsymbol{\psi}_i^\top \boldsymbol{\beta}$ $(i = 1, 2, 3)$ where $\boldsymbol{\psi}_i$ is a known 3D vector and $\boldsymbol{\beta} = [\sin^2(\theta), \sin(\theta)\cdot \cos(\theta), \cos^2(\theta)]^\top$. The norms of $\mathbf{d}_2$ and $\mathbf{d}_3$ satisfy $\|\mathbf{d}_2\| = \|\mathbf{d}_3\| = \sqrt{\boldsymbol{\mu}^\top \boldsymbol{\beta}}$ where $\boldsymbol{\mu}$ is known. Note that unlike the norm $\|\mathbf{d}_1\| \equiv 1$, $\|\mathbf{d}_2\|$ and $\|\mathbf{d}_3\|$ change with respect to the unknown angle θ. We normalize $\mathbf{d}_2$ and $\mathbf{d}_3$ by $\bar{\mathbf{d}}_2 = \mathbf{d}_2/\sqrt{\boldsymbol{\mu}^\top \boldsymbol{\beta}}$ and $\bar{\mathbf{d}}_3 = \mathbf{d}_3/\sqrt{\boldsymbol{\mu}^\top \boldsymbol{\beta}}$, respectively.

Based on the above orthogonal unit 3D directions $\{\mathbf{d}_1, \bar{\mathbf{d}}_2, \bar{\mathbf{d}}_3\}$, we parametrize the MF rotation $\mathbf{R}^{\mathcal{M}\to\mathcal{C}}$ from the MF $\mathcal{M}$ to the camera frame $\mathcal{C}$. As shown in Fig. 3, the 3D lines $\mathbf{L}_1$ and $\mathbf{L}_2$ are aligned to two axes of $\mathcal{M}$. Without loss of generality, we associate $\mathbf{d}_1$, $\bar{\mathbf{d}}_2$ and $\bar{\mathbf{d}}_3$ with the x-, y- and z-axes of $\mathcal{M}$, respectively. Accordingly, their coordinates in $\mathcal{M}$ are $\mathbf{d}_1^{\mathcal{M}} = [1, 0, 0]^\top$, $\bar{\mathbf{d}}_2^{\mathcal{M}} = [0, 1, 0]^\top$ and $\bar{\mathbf{d}}_3^{\mathcal{M}} = [0, 0, 1]^\top$. Based on the constraint that $\mathbf{R}^{\mathcal{M}\to\mathcal{C}}[\mathbf{d}_1^{\mathcal{M}}, \bar{\mathbf{d}}_2^{\mathcal{M}}, \bar{\mathbf{d}}_3^{\mathcal{M}}] = [\mathbf{d}_1, \bar{\mathbf{d}}_2, \bar{\mathbf{d}}_3]$, we obtain the MF rotation $\mathbf{R}^{\mathcal{M}\to\mathcal{C}}$ as

$$\mathbf{R}^{\mathcal{M}\to\mathcal{C}}(\theta) = \left[\underbrace{\begin{matrix} \boldsymbol{\omega}_{1,1}^\top \boldsymbol{\alpha} \\ \boldsymbol{\omega}_{1,2}^\top \boldsymbol{\alpha} \\ \boldsymbol{\omega}_{1,3}^\top \boldsymbol{\alpha} \end{matrix}}_{\mathbf{d}_1(\theta)} \; \underbrace{\begin{pmatrix} \boldsymbol{\omega}_{2,1}^\top \boldsymbol{\alpha} & \boldsymbol{\psi}_1^\top \boldsymbol{\beta} \\ \boldsymbol{\omega}_{2,2}^\top \boldsymbol{\alpha} & \boldsymbol{\psi}_2^\top \boldsymbol{\beta} \\ \boldsymbol{\omega}_{2,3}^\top \boldsymbol{\alpha} & \boldsymbol{\psi}_3^\top \boldsymbol{\beta} \end{pmatrix} \cdot \frac{1}{\sqrt{\boldsymbol{\mu}^\top \boldsymbol{\beta}}}}_{\bar{\mathbf{d}}_2(\theta), \bar{\mathbf{d}}_3(\theta)} \right]. \tag{3}$$

Eq. (3) shows that our MF rotation is composed of the 3D directions $\{\mathbf{d}_1, \bar{\mathbf{d}}_2, \bar{\mathbf{d}}_3\}$. In essence, these directions encode three DOF of the MF rotation. We use the image lines $\mathbf{l}_1$ and $\mathbf{l}_2$ to estimate its two DOF by Eqs. (1) and (2), and parametrize the MF rotation by the single parameter θ encoding its third DOF. Therefore, we reduce the rotation space as a 1D space, which speeds up our search (see Section 5.2). Moreover, our parametrization contributes to computing tight bounds of our cost function efficiently, which further accelerates our search (see Section 5.3).

4.2. Sampling Two Image Lines

In Section 4.1, to compute two DOF of the MF rotation, we use two sampled image lines corresponding to two different VPs. However, we do not have the prior knowledge regarding which two image lines among the extracted lines satisfy this assumption. In addition, the extracted lines may be corrupted by outliers that do not correspond to any VP. To solve these problems, we randomly sample two image lines S times to guarantee at least one sampling satisfying our assumption (called the "valid" sampling). We compute S following RANSAC [13] as $S = \frac{\log(1-c)}{\log(1-p)}$ where c is the confidence level, and p is the probability that two sampled lines correspond to two different VPs. We set c as 0.99 and set p as 15%, and thus $S \approx 28$.

We also propose a method to improve the efficiency and robustness of sampling. Fig. 4 shows that in numerous practical cases, at least one VP corresponds to a set of nearly parallel image lines [35]. We generate a histogram of the image line directions. If the bin with the highest cardinality corresponds to a sharp peak (see red bin in Fig. 4), we sample the first line from this bin and the second line from the remaining bins (otherwise we use the random sampling). Accordingly, we set p as 30% and $S \approx 13$. After S iterations, we hypothesize S MF rotations $\{\mathbf{R}_s^{\mathcal{M}\to\mathcal{C}}(\theta)\}_{s=1}^S$ (with the unknown third DOF θ) by Eq. (3). Note that we may generate more than one valid and several invalid hypotheses. We efficiently identify the optimal one in Section 5.

5. Searching for the Third DOF

In this section, we search for the third DOF of the MF rotation. We model the search as the inlier set maximization and solve it by BnB. Moreover, we propose an efficient method to compute tight bounds of our cost function.

5.1. Inlier Set Maximization

In Section 4.2, we hypothesized S MF rotations $\mathcal{R} = \{\mathbf{R}_s^{\mathcal{M}\to\mathcal{C}}(\theta)\}_{s=1}^S$ (with the unknown third DOF θ). To obtain the optimal MF rotation, two challenges exist: 1) among $\mathcal{R}$, how to identify the optimal hypothesis $\hat{\mathbf{R}}^{\mathcal{M}\to\mathcal{C}}(\theta)$; 2) for $\hat{\mathbf{R}}^{\mathcal{M}\to\mathcal{C}}(\theta)$, how to obtain its optimal parameter $\hat{\theta}$. We tackle these two challenges as an inlier set maximization problem [13]. Given a set of image lines, we aim at finding

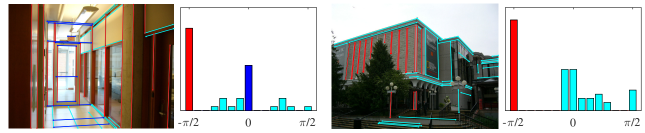

Figure 4. Image lines and histograms of their directions. Several nearly parallel image lines, which are associated with a bin with high cardinality, correspond to the same VP.

the optimal MF rotation $\hat{\mathbf{R}}^{\mathcal{M}\to\mathcal{C}}(\hat{\theta})$ maximizing the number of inlier image lines whose associated 3D lines are aligned to the MF axes.

We present the mathematical formulation as follows. We denote the m-th column of the MF rotation $\mathbf{R}^{\mathcal{M}\to\mathcal{C}}(\theta)$ (i.e. the m-th MF axis) by $[\mathbf{R}^{\mathcal{M}\to\mathcal{C}}(\theta)]^m$ $(m = 1, 2, 3)$. In the noise-free case, for an inlier image line $\mathbf{l}_k$, its associated unit projection plane normal $\mathbf{n}_k$ is orthogonal to a MF axis $[\mathbf{R}^{\mathcal{M}\to\mathcal{C}}(\theta)]^m$ (see Fig. 3). Under the presence of noise, we define the residual as

$$\epsilon_k^m(\theta) = \left|\mathbf{n}_k^\top [\mathbf{R}^{\mathcal{M}\to\mathcal{C}}(\theta)]^m\right|. \tag{4}$$

Accordingly, we define the line $\mathbf{l}_k$ as an inlier if its residual $\epsilon_k^m(\theta)$ is smaller than the threshold $\tau = \cos\left(\frac{\pi}{2} - \frac{\pi}{90}\right)$, i.e. the angle error is smaller than $2°$. For the s-th hypothesized MF rotation $\mathbf{R}_s^{\mathcal{M}\to\mathcal{C}}(\theta) \in \mathcal{R}$, we compute its number of inliers $N_s(\theta)$ among K extracted lines as

$$N_s(\theta) = \sum_{k=1}^{K} \sum_{m=1}^{3} f\left(\epsilon_k^m(\theta)\right), \tag{5}$$

where $f\left(\epsilon_k^m(\theta)\right)$ represents the inlier function, i.e.

$$f\left(\epsilon_k^m(\theta)\right) = \begin{cases} 1, & \text{if } \epsilon_k^m(\theta) \leqslant \tau; \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Let $\mathring{N}_s(\theta)$ denote the unknown maximum of $N_s(\theta)$. We aim at finding 1) the optimal hypothesis $\hat{\mathbf{R}}^{\mathcal{M}\to\mathcal{C}}(\theta) \in \mathcal{R}$ (i.e. $\hat{s} \in \{1, \cdots, S\}$), and 2) its optimal parameter $\hat{\theta} \in [0, \pi]$ to achieve the global maximum of $\{\mathring{N}_s(\theta)\}_{s=1}^S$ as

$$\hat{\mathbf{R}}^{\mathcal{M}\to\mathcal{C}}(\hat{\theta}) = \arg\max_{s,\theta} \{\mathring{N}_s(\theta)\}_{s=1}^S. \tag{7}$$

Compared with the algebraic error used by [25, 35], our residual $\epsilon_k^m(\theta)$ in Eq. (5) corresponds to the geometric error that is known to be more meaningful [17]. Note that while we simplify the MF rotation by computing its two DOF, solving Eq. (7) is still relatively challenging. First, both the line clusters and the parameter θ are unknown, which constitutes a non-convex problem [10]. Second, we reduce the search space at the cost of hypothesizing several invalid (or sub-optimal) MF rotations. In theory, testing each hypothesis and exhaustively searching over the interval $[0, \pi]$ of θ could find the optimal solution, but it is not practical due to high complexity. In addition, the gradient descent method [7] is sensitive to the initialization of the hypothesis and parameter, and may get stuck into a local optimum.

5.2. Search Based on BnB

We employ BnB to solve Eq. (7) by searching for the optimal hypothesized MF rotation and its optimal parameter. BnB is a popular method to provide the globally optimal solution maximizing the number of inliers. It has been used for various applications such as camera pose estimation [10, 16, 27] and point set registration [8, 9, 34]. It divides the search space into several sub-spaces, and computes the upper and lower bounds of the cost function for each sub-space. A sub-space is discarded if its associated bounds prove it does not contain the optimal solution (tight bounds thus accelerate discarding sub-spaces). Remaining sub-spaces are further divided and discarded until the optimal solution is found. In our context, the search space represents the interval $[0, \pi]$ of the parameter θ, and the cost function represents the number of inliers in Eq. (7).

Fig. 5(a) shows that each hypothesized MF rotation corresponds to a binary tree. We continuously divide the original interval $[0, \pi]$ of θ and treat the sub-intervals as the child nodes. For the node Θ of the s-th tree, we assign it with the number of inliers $N_s(\Theta)$ (see Eq. (5)). Note that Θ represents an interval of θ rather than a number. Therefore, we compute the bounds rather than the specific value of $N_s(\Theta)$. We denote the lower and upper bounds of $N_s(\Theta)$ by $\underline{N_s}(\Theta)$ and $\overline{N_s}(\Theta)$ respectively, and introduce how to compute them in Section 5.3. We adopt the best-first search strategy [34], i.e. the interval whose upper bound is high has a high priority. In the following, we take some trees for example to explain how we search for the optimal hypothesis and its optimal parameter.

As shown in Fig. 5(a), we use the red tree ("r") to show our search of the optimal parameter. $\overline{N_{\mathrm{r}}}(\Theta_{\text{III-2}})$ is lower than $\underline{N_{\mathrm{r}}}(\Theta_{\text{III-1}})$, proving that the optimal solution of θ is not within the interval $\Theta_{\text{III-2}}$. Therefore, we discard $\Theta_{\text{III-2}}$. To accelerate the search, we also compute the midpoint of the interval $\Theta_{\text{III-1}}$, which is denoted by $\check{\Theta}_{\text{III-1}}$. While $\overline{N_{\mathrm{r}}}(\Theta_{\text{II-2}})$ is higher than $\underline{N_{\mathrm{r}}}(\Theta_{\text{III-1}})$, we still discard the node $\Theta_{\text{II-2}}$ since $\overline{N_{\mathrm{r}}}(\Theta_{\text{II-2}})$ is lower than $N_{\mathrm{r}}(\check{\Theta}_{\text{III-1}})$. In addition, as shown in Fig. 5(b), we use the green tree ("g") and the blue tree ("b") to show our search of the optimal hypothesis. For the remaining intervals of each tree, we compute the bounds of their associated numbers of inliers (see Section 5.3). Then we compute the maximum of these upper bounds and the minimum of these lower bounds, which are called the *local* maximum and minimum respectively. The local maximum $\overline{N_{\mathrm{b}}}(\Theta_{\text{IV-1}})$ is lower than the local minimum $\underline{N_{\mathrm{g}}}(\Theta_{\text{IV-2}})$, proving that the hypothesis $\mathbf{R}_{\mathrm{b}}^{\mathcal{M}\to\mathcal{C}}(\theta)$ is less accurate than the hypothesis $\mathbf{R}_{\mathrm{g}}^{\mathcal{M}\to\mathcal{C}}(\theta)$. Therefore, we discard the blue tree and only search over the red and green trees. In addition, for all these trees, we compute the maximum of their local maxima, which is called the *global* maximum. As shown in Fig. 5(c), the interval $\Theta_{\text{V-1}}$ of the red tree satisfies our stopping criterion. Specifically, the number of inliers computed by its midpoint $\check{\Theta}_{\text{V-1}}$ equals to the global maximum. We treat $\mathbf{R}_{\mathrm{r}}^{\mathcal{M}\to\mathcal{C}}(\check{\Theta}_{\text{V-1}})$ as the optimal MF rotation since it maximizes the number of inliers. We do not consider the green tree since its local maximum is not higher than (at most equals to) the global maximum.

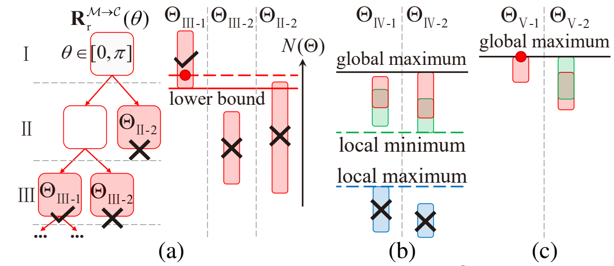

Figure 5. (a) Dividing the parameter interval Θ associated with the number of inliers $N(\Theta)$. {I,II,III...} denote the numbers of divisions. (b) Discarding the invalid (or sub-optimal) hypothesized MF rotation. (c) Stopping criterion. The red point in (a) or (c) denotes the number of inliers computed by an interval midpoint.

Overall, we search for the third DOF of the MF rotation by fixing its two DOF computed by sampling, and thus obtain the quasi-globally optimal MF rotation efficiently. Moreover, for the cases with "fake" MF(s) defined by outliers, the pure search-based methods fail, and the pure sampling-based methods may not be accurate. In contrast, our approach provides accurate VPs thanks to our hybrid strategy, as will be shown in the experiments.

5.3. Bounds of the Number of Inliers

We propose a novel method to efficiently compute tight bounds of the number of inliers, i.e. $\underline{N_s}(\Theta)$ and $\overline{N_s}(\Theta)$ used in Section 5.2. We begin with computing the bounds of the residual $\epsilon_k^m(\theta)$ in Eq. (4). By substituting Eq. (3) into Eq. (4), we rewrite $\epsilon_k^m(\theta)$ as three different types:

$$\epsilon_k^m(\theta) = \begin{cases} \big|\underbrace{\boldsymbol{\xi}^\top \boldsymbol{\alpha}}_{\mathsf{L}^{\xi}}\big|, & m=1 \ (x\text{-axis of MF}); \\ \big|\underbrace{\boldsymbol{\zeta}^\top \boldsymbol{\alpha}}_{\mathsf{L}^{\zeta}} / \underbrace{\sqrt{\boldsymbol{\mu}^\top \boldsymbol{\beta}}}_{\sqrt{\mathsf{Q}^{\mu}}}\big|, & m=2 \ (y\text{-axis of MF}); \\ \big|\underbrace{\boldsymbol{\varphi}^\top \boldsymbol{\beta}}_{\mathsf{Q}^{\varphi}} / \underbrace{\sqrt{\boldsymbol{\mu}^\top \boldsymbol{\beta}}}_{\sqrt{\mathsf{Q}^{\mu}}}\big|, & m=3 \ (z\text{-axis of MF}). \end{cases} \tag{8}$$

where the known $\boldsymbol{\xi}$, $\boldsymbol{\zeta}$ and $\boldsymbol{\varphi}$ are computed by the known $\{\boldsymbol{\omega}_{1,i}\}_{i=1}^3$, $\{\boldsymbol{\omega}_{2,i}\}_{i=1}^3$ and $\{\boldsymbol{\psi}_i\}_{i=1}^3$ in Eq. (3), respectively; $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ composed of $\sin(\theta)$ and $\cos(\theta)$, as well as the known $\boldsymbol{\mu}$, are defined in Section 4.1. Eq. (8) shows that the elements of $\epsilon_k^m(\theta)$ satisfy two forms: 1) $a \cdot \sin(\theta) + b \cdot \cos(\theta)$ that is called the "linear trigonometric polynomial" and denoted by "L"; 2) $c \cdot \sin^2(\theta) + d \cdot \sin(\theta) \cdot \cos(\theta) + e \cdot \cos^2(\theta)$ that is called the "quadratic trigonometric polynomial" and denoted by "Q". We reformulate computing the bounds of the residual $\epsilon_k^m(\theta)$ as computing the bounds of its elements, i.e. the trigonometric polynomials L and Q.

To compute the bounds of L, we transform it as

$$\mathsf{L} = a \cdot \sin(\theta) + b \cdot \cos(\theta) = u_1 \cdot \sin(\theta + v_1), \tag{9}$$

where $u_1 = \sqrt{a^2 + b^2}$ and $v_1 = \arctan(b/a)$. To compute the bounds of Q, we transform it as

$$\begin{aligned} \mathsf{Q} &= c \cdot \sin^2(\theta) + d \cdot \sin(\theta) \cdot \cos(\theta) + e \cdot \cos^2(\theta) \\ &= (d/2) \cdot \sin(2\theta) + \big((e-c)/2\big) \cdot \cos(2\theta) + (c+e)/2 \\ &= u_2 \cdot \sin(2\theta + v_2) + w_2, \end{aligned} \tag{10}$$

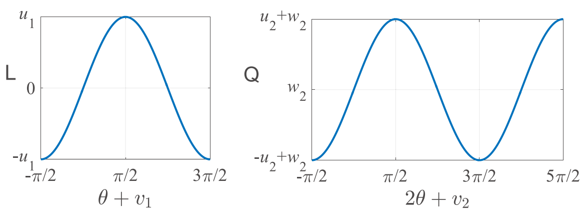

Figure 6. Curves of the trigonometric polynomials expressed by the sine functions: (left) linear trigonometric polynomial L in Eq. (9); (right) quadratic trigonometric polynomial Q in Eq. (10).

where $u_2 = \sqrt{d^2+(e-c)^2}/2$, $v_2 = \arctan\left((e-c)/d\right)$ and $w_2 = (c+e)/2$. We thus express L and Q by $\sin(\theta+v_1)$ and $\sin(2\theta+v_2)$, respectively. The sign ("+" or "-") of u_1 and u_2 depends on the sign of a and d, respectively. Without loss of generality, we introduce the case that $u_1 > 0$ and $u_2 > 0$. Recall that the unknown parameter θ is within $[0, \pi]$, and v_1 in Eq. (9) and v_2 in Eq. (10) are both within $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Accordingly, $\theta+v_1$ of L is within $[-\frac{\pi}{2}, \frac{3\pi}{2}]$, and $2\theta+v_2$ of Q is within $[-\frac{\pi}{2}, \frac{5\pi}{2}]$. Fig. 6 shows the curves of L and Q, and we use it to compute the bounds of L and Q as follows.

Given an interval of θ denoted by $\Theta = [\underline{\theta}, \overline{\theta}]$, we first compute 1) the domain of L, i.e. $\mathcal{D}_1 = [\underline{\theta}+v_1, \overline{\theta}+v_1] \in [-\frac{\pi}{2}, \frac{3\pi}{2}]$ (denoted by $[\underline{\mathcal{D}_1}, \overline{\mathcal{D}_1}]$), and 2) the domain of Q, i.e. $\mathcal{D}_2 = [2\underline{\theta}+v_2, 2\overline{\theta}+v_2] \in [-\frac{\pi}{2}, \frac{5\pi}{2}]$ (denoted by $[\underline{\mathcal{D}_2}, \overline{\mathcal{D}_2}]$). Then we efficiently obtain the ranges, i.e. the *strict* bounds of L and Q. Specifically, as shown in Fig. 6(left), we obtain the strict bounds of L by judging whether the stationary point $\frac{\pi}{2}$ is within its domain $\mathcal{D}_1$ as

$$\begin{cases} \underline{\mathsf{L}} = u_1 \cdot \min\left(\sin(\underline{\mathcal{D}_1}), \sin(\overline{\mathcal{D}_1})\right); \\ \overline{\mathsf{L}} = \begin{cases} u_1 & \text{if } \frac{\pi}{2} \in \mathcal{D}_1; \\ u_1 \cdot \max\left(\sin(\underline{\mathcal{D}_1}), \sin(\overline{\mathcal{D}_1})\right) & \text{otherwise.} \end{cases} \end{cases} \tag{11}$$

As shown in Fig. 6(right), we obtain the strict bounds of Q by judging whether the stationary points $\{\frac{\pi}{2}, \frac{3\pi}{2}\}$ are within its domain $\mathcal{D}_2$ as

$$\begin{cases} \begin{cases} \underline{\mathsf{Q}} = -u_2 + w_2 \\ \overline{\mathsf{Q}} = u_2 + w_2 \end{cases} & \text{if } \begin{cases} \frac{\pi}{2} \in \mathcal{D}_2 \\ \frac{3\pi}{2} \in \mathcal{D}_2 \end{cases}; \\ \begin{cases} \underline{\mathsf{Q}} = u_2 \cdot \min\left(\sin(\underline{\mathcal{D}_2}), \sin(\overline{\mathcal{D}_2})\right) + w_2 \\ \overline{\mathsf{Q}} = u_2 + w_2 \end{cases} & \text{if } \begin{cases} \frac{\pi}{2} \in \mathcal{D}_2 \\ \frac{3\pi}{2} \notin \mathcal{D}_2 \end{cases}; \\ \begin{cases} \underline{\mathsf{Q}} = -u_2 + w_2 \\ \overline{\mathsf{Q}} = u_2 \cdot \max\left(\sin(\underline{\mathcal{D}_2}), \sin(\overline{\mathcal{D}_2})\right) + w_2 \end{cases} & \text{if } \begin{cases} \frac{\pi}{2} \notin \mathcal{D}_2 \\ \frac{3\pi}{2} \in \mathcal{D}_2 \end{cases}; \\ \begin{cases} \underline{\mathsf{Q}} = u_2 \cdot \min\left(\sin(\underline{\mathcal{D}_2}), \sin(\overline{\mathcal{D}_2})\right) + w_2 \\ \overline{\mathsf{Q}} = u_2 \cdot \max\left(\sin(\underline{\mathcal{D}_2}), \sin(\overline{\mathcal{D}_2})\right) + w_2 \end{cases} & \text{if } \begin{cases} \frac{\pi}{2} \notin \mathcal{D}_2 \\ \frac{3\pi}{2} \notin \mathcal{D}_2 \end{cases}. \end{cases} \tag{12}$$

Based on the bounds of L and Q obtained by Eqs. (11) and (12), we compute the bounds of the residual $\epsilon_k^m(\Theta)$ in Eq. (8). Specifically, for $\epsilon_k^1(\Theta)$, we use the bounds of L^ξ; for $\epsilon_k^2(\Theta)$, we use the bounds of $\{\mathsf{L}^\zeta, \mathsf{Q}^\mu\}$; for $\epsilon_k^3(\Theta)$, we use the bounds of $\{\mathsf{Q}^\varphi, \mathsf{Q}^\mu\}$. Our residual bound computation exploits the absolute value operation of a single interval[1] and/or the division operation between two intervals[2] [26]. Compared with [5, 6], we obtain tighter bounds of the residuals in a more efficient way. The reasons are: 1) our bounds of the trigonometric polynomials are strict (without any relaxation) and computed by simple algebraic operations; 2) we relax our residual bounds (by the division operation) only once for $\epsilon_k^2(\Theta)$ and $\epsilon_k^3(\Theta)$, and do not relax our residual bounds for $\epsilon_k^1(\Theta)$.

[1] $\left|[x,y]\right| = \left[\min(|x|,|y|), \max(|x|,|y|)\right]$, if $x \cdot y > 0$; $\left|[x,y]\right| = \left[0, \max(|x|,|y|)\right]$, if $x \cdot y \leqslant 0$ where x and y represent the lower and upper bounds of an interval, respectively.

Algorithm 1: MF rotation computation

Input: extracted lines $\{\mathbf{l}_k\}$ in a calibrated image.

Step one — estimating two DOF by sampling:

Sample two image lines S times (see Section 4.2);

for each pair of the sampled lines **do**

  Parametrize a hypothesized MF rotation using the rotation angle θ by Eq. (3);

end

Step two — searching for the third DOF (θ):

repeat

  for each hypothesized MF rotation $\mathbf{R}_s^{\mathcal{M}\to\mathcal{C}}(\theta)$ **do**

    Divide a remaining interval of θ (best-first);

    for each sub-interval **do**

      for each image line $\mathbf{l}_k$ **do**

        Compute the bounds of the polynomials by Eqs. (11) and (12);

      end

      Compute the bounds of the number of inliers by Eq. (13);

    end

    Discard sub-optimal intervals of θ (see Fig. 5);

  end

  Discard invalid MF rotation hypotheses (see Fig. 5);

until satisfy our stopping criterion (see Section 5.2).

Output: the optimal MF rotation $\hat{\mathbf{R}}^{\mathcal{M}\to\mathcal{C}}(\hat{\theta})$.

Based on our tight bounds of the residuals and Eq. (5), we finally compute the bounds of the number of inliers $\underline{N_s}(\Theta)$ and $\overline{N_s}(\Theta)$ used in Section 5.2 as

$$[\underline{N_s}(\Theta), \overline{N_s}(\Theta)] = \sum_{k=1}^{K}\sum_{m=1}^{3}\left[\ \underline{f}\left(\epsilon_k^m(\Theta)\right), \overline{f}\left(\epsilon_k^m(\Theta)\right)\right], \tag{13}$$

where $[\underline{f}, \overline{f}]$ has three types of values based on Eq. (6): 1) $[\underline{f}, \overline{f}] = [1, 1]$ if $\overline{\epsilon}(\Theta) \leqslant \tau$, i.e. the image line $\mathbf{l}_k$ is definitely an inlier; 2) $[\underline{f}, \overline{f}] = [0, 0]$ if $\tau \leqslant \underline{\epsilon}(\Theta)$, i.e. $\mathbf{l}_k$ is definitely an outlier; 3) $[\underline{f}, \overline{f}] = [0, 1]$ if $\underline{\epsilon}(\Theta) < \tau < \overline{\epsilon}(\Theta)$, i.e. whether $\mathbf{l}_k$ is an inlier is uncertain.

Our full MF rotation computation method for VP estimation is described in Algorithm 1.

6. Experiments

We have conducted experiments on both synthetic and real-world images. Source codes and supplementary materials are available on our project website[3]. We list the state-

[2] $[x,y]/[x',y'] = \left[\min(\frac{x}{x'}, \frac{x}{y'}, \frac{y}{x'}, \frac{y}{y'}), \max(\frac{x}{x'}, \frac{x}{y'}, \frac{y}{x'}, \frac{y}{y'})\right]$.

[3] https://sites.google.com/view/haoangli/projects/iccv-vp

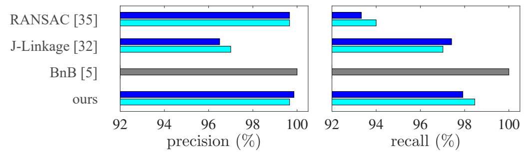

Figure 7. Accuracy comparison in an outlier-free case: (left) precision; (right) recall. We show mean and median in blue and cyan, respectively (we run **BnB** [5] once and show its results in gray).

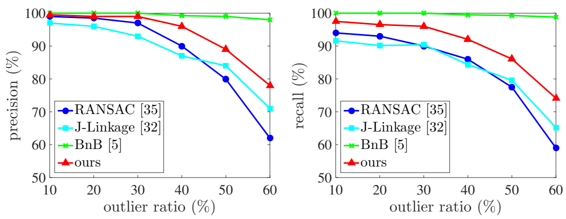

Figure 8. Accuracy comparison with respect to the outlier ratio: average precision (left) and recall (right).

of-the-art methods introduced in Section 2 for comparison:

- **RANSAC** that retrieves the optimal MF rotation hypothesized by three sampled image lines [35].
- **J-Linkage** that clusters image lines by the descriptor similarity [31].
- **BnB** that searches for the optimal MF rotation over the rotation space [5].

We compute the number of iterations of **RANSAC** based on the valid sampling probability 0.1 (smaller than our probability 0.15 in Section 4.2 due to its stronger assumption on the configuration of three 3D lines). We set the times of sampling of **J-Linkage** as 500 (recommended by [31]). We also conduct experiments by adjusting the time budgets of various methods. For a more unbiased comparison of the original accuracy, we report the raw results without iterative refinement. All the methods are implemented in MATLAB and tested on an Intel Core i7 CPU with 3.60 GHz.

Evaluation criteria. We evaluate the VP estimation accuracy in terms of the precision and recall of the image line clustering [19]. The precision is defined by $N^{\mathrm{c}}/(N^{\mathrm{c}}+N^{\mathrm{w}})$ where N^{c} and N^{w} are the numbers of correctly and wrongly clustered lines, respectively. The recall is defined by $N^{\mathrm{c}}/(N^{\mathrm{c}}+N^{\mathrm{m}})$ where N^{m} is the number of missing inliers. For the efficiency evaluation, in addition to comparing the absolute run time of various methods, we compare **BnB** with our method in terms of the number of interval divisions [5, 20].

6.1. Synthetic Images

We synthesize two or three sets of parallel 3D lines. Each two lines from different sets are mutually orthogonal. We project them by a synthetic camera, obtaining two or three sets of image lines. We perturb the endpoints of these image lines by a zero-mean Gaussian noise whose standard deviation is σ pixels. We generate outlier image lines by randomizing their positions and directions. We aim at clustering these image lines by unknown-but-sought VPs.

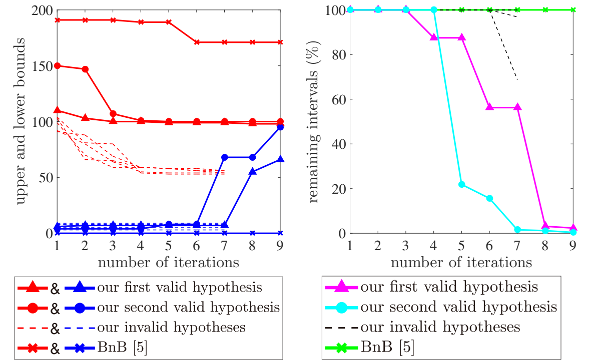

Figure 9. Efficiency comparison in terms of the number of iterations: (left) evolution of the upper and lower bounds of the number of inliers (we report the local maxima and minima of hypotheses); (right) evolution of the proportion of remaining intervals.

Table 1. Computational time comparison

	RANSAC [35]	**J-Linkage** [31]	**BnB** [5]	**ours**
time (s)	0.011	0.187	3.501	0.276

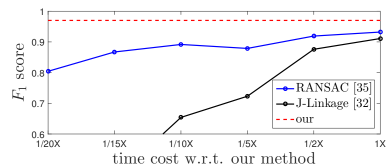

Figure 10. F_1-score comparison by fixing the time budget for our method and changing it for **RANSAC** [35] and **J-Linkage** [31].

Accuracy evaluation. Due to the sampling uncertainty, we conduct 1000 trials for **RANSAC**, **J-Linkage** and our method. Fig. 7 shows a representative comparison with 3 VPs and 150 inlier lines. The noise level is $\sigma = 5$. **RANSAC** fails to completely retrieve inliers due to the effect of noise. **J-Linkage** mistakenly clusters several lines since its line descriptors may be contaminated by noise. **BnB** successfully retrieves all the inliers. Our method achieves high precision and recall thanks to our hybrid method for MF rotation computation. In addition, Fig. 8 shows the comparison under the presence of outliers. Given 150 lines, we fix the noise level as $\sigma = 3$ and vary the outlier ratio from 10% to 60%. High outlier ratios increase the difficulty of valid sampling. **RANSAC** and **J-Linkage** obtain unsatisfactory results for more than 30% outliers. In contrast, our method is more robust since it requires less samples. Note that **BnB** becomes unstable for high outlier ratios since it may retrieve a "fake" MF defined by numerous outliers, which will be fully analyzed in Section 6.2.

Efficiency evaluation. We report the result of a representative test with 2 VPs and 100 inliers. The noise level is $\sigma = 4$. We first compare **BnB** with our method in terms of the number of iterations. We hypothesize 13 MF rotations by 2 valid and 11 invalid samplings (the validity is identified by the ground truth cluster labels). Fig. 9(left) shows the evolution of the bounds of the number of inliers. Without loss of generality, we randomly select 5 invalid hy-

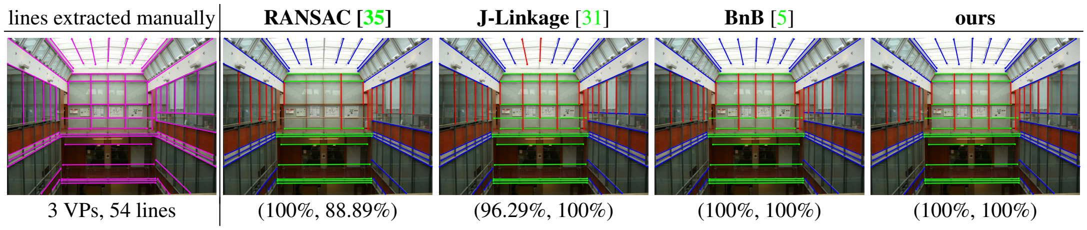

Figure 11. VP estimation using the manually extracted image lines on the York Urban Database [12]. Different image line clusters are shown in respective colors. The pair of numbers below each image represents the precision and recall of image line clustering.

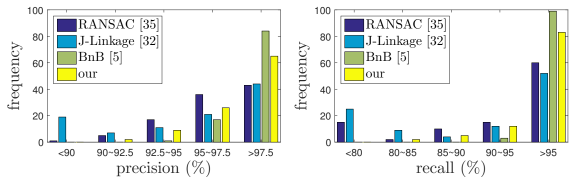

Figure 12. Accuracy comparison on 102 images of YUD [12]: precision (left) and recall (right).

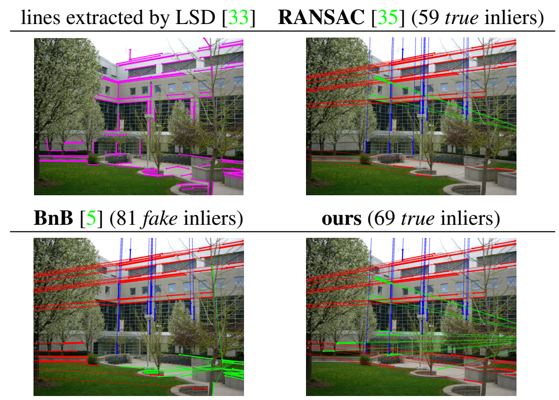

Figure 13. VP estimation using the image lines corrupted by outliers. The connection between a clustered line and its corresponding VP is shown in dotted line (the angle between its projection plane normal and its corresponding MF axis is $90 \pm 2°$).

potheses to report. At the 7-th iteration, for all the invalid hypotheses, their local maxima of upper bounds are smaller than the local minimum of lower bounds of our second valid hypothesis, and thus they are discarded. At the 9-th iteration, an interval associated with our second valid hypothesis satisfies our stopping criterion (see Section 5.2). Therefore, our method takes only 9 iterations to obtain all the inliers. In contrast, **BnB** requires 318 iterations.

Fig. 9(right) shows the evolution of the proportion of remaining intervals (with respect to the length for our method or volume for **BnB**). At convergence, our method discards most intervals thanks to our tight bounds, but **BnB** still keeps numerous intervals. In addition, we conduct 500 independent trials using the above data and present the average run time of various methods in Table 1. **RANSAC** is efficient at the cost of sacrificing accuracy. Our method is significantly faster than **BnB** and is similar to **J-Linkage**. In addition, we fix the time cost t of our method as about 0.2s, and vary the time budget of **RANSAC** and **J-Linkage** from $\frac{t}{20}$ to t under the outlier ratio 40%. As shown in Fig. 10, while **RANSAC** and **J-Linkage** improve their accuracy as their time budgets increase, our method achieves the highest F_1-score [7] given the same time budget t.

6.2. Real-world Images

To evaluate various methods on real-world images, we conduct experiments on the York Urban Database [12]. It is composed of 102 calibrated images of 640×480 pixels, and each image contains a set of manually extracted lines. These lines are outlier-free and correspond to 2 or 3 VPs. Fig. 11 shows a representative comparison, and Fig. 12 reports the precision and recall of various method on all the images. **RANSAC** and **J-Linkage** fail to guarantee the precision and recall simultaneously. **BnB** stably obtains all the inliers. Our method is less sensitive to noise than **RANSAC** and **J-Linkage**, and achieves similar accuracy to **BnB**. In addition, Fig. 13 shows the clustering result of a set of image lines automatically extracted by LSD [33]. Numerous outliers and partial inliers define a "fake" MF that is associated with more image lines than the ground truth MF. **BnB** retrieves the fake MF and mistakenly clusters numerous lines. In contrast, **RANSAC** and our method obtain correct clusters by sampling three and two image lines associated with the ground truth MF, respectively. Moreover, our method obtains a larger number of inliers than **RANSAC** thanks to our search of the third DOF of the MF rotation.

7. Conclusion

We proposed a hybrid MF rotation computation approach to estimate VPs in Manhattan world. We first compute two DOF by two sampled image lines and then search for the third DOF by BnB. Our sampling speeds up our search by reducing the search space and simplifying the bound computation. Our search is not sensitive to noise and achieves quasi-global optimality. Experiments on synthetic and real-world images showed that our approach outperforms state-of-the-art methods in terms of accuracy and/or efficiency. In the future, we will focus on computing one DOF by sampling and two DOF by search.

Acknowledgments. This work is supported in part by the Natural Science Foundation of China under Grant U1613218, in part by the Hong Kong ITC under Grant ITS/448/16FP, in part by the VC Fund 4930745 of the CUHK T Stone Robotics Institute, and in part by the National Research Foundation of Korea (Grant NRF-2017R1C1B5077030).

References

[1] Matthew Antone and Seth Teller. Automatic recovery of relative camera rotations for urban scenes. In *CVPR*, 2000. 2

[2] Stephen Barnard. Interpreting perspective images. *Artificial Intelligence*, 1983. 2

[3] Jean-Charles Bazin, Cedric Demonceaux, Pascal Vasseur, and Inso Kweon. Rotation estimation and vanishing point extraction by omnidirectional vision in urban environment. *IJRR*, 2012. 2

[4] Jean-Charles Bazin and Marc Pollefeys. 3-line RANSAC for orthogonal vanishing point detection. In *IROS*, 2012. 1, 2

[5] Jean-Charles Bazin, Yongduek Seo, Cedric Demonceaux, Pascal Vasseur, Katsushi Ikeuchi, Inso Kweon, and Marc Pollefeys. Globally optimal line clustering and vanishing point estimation in Manhattan world. In *CVPR*, 2012. 1, 2, 6, 7, 8

[6] Jean-Charles Bazin, Yongduek Seo, and Marc Pollefeys. Globally optimal consensus set maximization through rotation search. In *ACCV*, 2012. 1, 2, 6

[7] Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. 4, 8

[8] Zhipeng Cai, Tat-Jun Chin, Alvaro Parra Bustos, and Konrad Schindler. Practical optimal registration of terrestrial LiDAR scan pairs. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019. 5

[9] Dylan Campbell and Lars Petersson. GOGMA: Globally-optimal gaussian mixture alignment. In *CVPR*, 2016. 5

[10] Dylan Campbell, Lars Petersson, Laurent Kneip, and Hongdong Li. Globally-optimal inlier set maximisation for simultaneous camera pose and feature correspondence. In *ICCV*, 2017. 4, 5

[11] James Coughlan and Alan Yuille. Manhattan world: Compass direction from a single image by Bayesian inference. In *ICCV*, 1999. 1

[12] Patrick Denis, James Elder, and Francisco Estrada. Efficient edge-based methods for estimating Manhattan frames in urban imagery. In *ECCV*, 2008. 2, 8

[13] Martin Fischler and Robert Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981. 1, 2, 4

[14] Alex Flint, David Murray, and Ian Reid. Manhattan scene understanding using monocular, stereo, and 3D features. In *ICCV*, 2011. 1

[15] Yuan Gao and Alan Yuille. Exploiting symmetry and/or Manhattan properties for 3D object structure estimation from single and multiple images. In *CVPR*, 2017. 1

[16] Richard Hartley and Fredrik Kahl. Global optimization through rotation space search. *IJCV*, 2009. 5

[17] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2003. 3, 4

[18] Kyungdon Joo, Tae-Hyun Oh, Junsik Kim, and In So Kweon. Robust and globally optimal Manhattan frame estimation in near real time. *TPAMI*, 2019. 2

[19] Manh Cuong Le, Son Lam Phung, and Abdesselam Bouzerdoum. Lane detection in unstructured environments for autonomous navigation systems. In *ACCV*, 2015. 7

[20] Hongdong Li. Consensus set maximization with guaranteed global optimality for robust geometry estimation. In *ICCV*, 2009. 1, 3, 7

[21] Haoang Li, Yazhou Xing, Ji Zhao, Jean-Charles Bazin, Zhe Liu, and Yun-Hui Liu. Leveraging structural regularity of Atlanta world for monocular SLAM. In *ICRA*, 2019. 1

[22] Haoang Li, Jian Yao, Jean-Charles Bazin, Xiaohu Lu, Yazhou Xing, and Kang Liu. A monocular SLAM system leveraging structural regularity in Manhattan world. In *ICRA*, 2018. 1

[23] Haoang Li, Ji Zhao, Jean-Charles Bazin, Wen Chen, Kai Chen, and Yun-Hui Liu. Line-based absolute and relative camera pose estimation in structured environments. In *IROS*, 2019. 1

[24] Xiaohu Lu, Jian Yao, Haoang Li, and Yahui Liu. 2-line exhaustive searching for real-time vanishing point estimation in Manhattan world. In *WACV*, 2017. 1

[25] Faraz Mirzaei and Stergios Roumeliotis. Optimal estimation of vanishing points in a Manhattan world. In *ICCV*, 2011. 1, 2, 4

[26] Ramon Moore, Baker Kearfott, and Michael Cloud. *Introduction to Interval Analysis*. Society for Industrial and Applied Mathematics, 2009. 6

[27] Carl Olsson, Fredrik Kahl, and Magnus Oskarsson. Branch-and-bound methods for Euclidean registration problems. *TPAMI*, 2009. 5

[28] Long Quan and Roger Mohr. Determining perspective structures using hierarchical Hough transform. *PRL*, 1989. 2

[29] Sudipta Sinha, Drew Steedly, Richard Szeliski, Maneesh Agrawala, and Marc Pollefeys. Interactive 3D architectural modeling from unordered photo collections. *TOG*, 2008. 2

[30] Julian Straub, Oren Freifeld, Guy Rosman, John Leonard, and John Fisher. The Manhattan frame model—Manhattan world inference in the space of surface normals. *TPAMI*, 2017. 1

[31] Jean-Philippe Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *ICCV*, 2009. 2, 7, 8

[32] Roberto Toldo and Andrea Fusiello. Robust multiple structures estimation with J-Linkage. In *ECCV*, 2008. 2

[33] Rafael Grompone von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, and Gregory Randall. LSD: A fast line segment detector with a false detection control. *TPAMI*, 2010. 8

[34] Jiaolong Yang, Hongdong Li, Dylan Campbell, and Yunde Jia. Go-ICP: A globally optimal solution to 3D ICP point-set registration. *TPAMI*, 2016. 5

[35] Lilian Zhang, Huimin Lu, Xiaoping Hu, and Reinhard Koch. Vanishing point estimation and line classification in a Manhattan world with a unifying camera model. *IJCV*, 2016. 1, 2, 3, 4, 7, 8

[36] Marco Zuliani, Charles Kenney, and Manjunath. The multi-RANSAC algorithm and its application to detect planar homographies. In *ICIP*, 2005. 2