Globally Optimal and Efficient Manhattan Frame Estimation by Delimiting Rotation Search Space

Wuwei Ge[2], Yu Song [*,1,2], Baichao Zhang[2], and Zehua Dong[2]

[1]Beijing Jiaotong University   [2]Zongmu Tech, Beijing, China

Abstract

A typical man-made structure can be abstracted as the Manhattan world assumption, in which notion is further represented as a Manhattan Frame (MF) defined by three orthogonal axes. The problem of MF estimation can be formulated as the solution of the rotation between the MF and the camera frame (called the "MF rotation"). However, the whole rotation space is quite redundant for solving the MF rotation, which is one of the main factors that disturb the computational efficiency of those methods associated with a rotation space search. This paper proves that the volume of the space that just contains all MF rotations (called the "MFR space") is only 1 / 24 of that of the whole rotation space, and then an exact MFR space is delimited from the rotation space. Searching in the delimited MFR space, the MF estimation solved by a branch-and-bound (BnB) framework guarantees stability and efficiency simultaneously. Furthermore, the general rotation problems associated with a rotation space search are solved more efficiently. Experiments on synthetic and real datasets have successfully confirmed the validity of our approach.

1. Introduction

Most man-made environments generally exhibit particular regularity like parallelism and orthogonality, which can be represented by a set of parallel and orthogonal structural lines. These structural forms are commonly abstracted as Manhattan world (MW) [1] assumption that consists of three mutually orthogonal directions corresponding to three orthogonal vanishing points (VP) in image plane of camera. By virtue of its simplicity, MW assumption has been successfully applied to many higher-level computer vision tasks, such as calibration of camera parameters [2, 3], scene understanding [7, 8, 9] and SLAM [10, 11, 12].

Since the introduction of the notion of Manhattan Frame (MF) [20], which is defined by three orthogonal axes corresponding to the three directions of MW, a considerable amount of work has been devoted to the stability and efficiency of MF estimation. However it is still challenging to satisfy both of the two properties. By formulating the problem of MF estimation as the solution of the rotation between the MF and the camera frame (called the "MF rotation"), state-of-the-art methods [15, 16] exhaustively search in the whole rotation space to meet a globally optimal solution in terms of maximizing the number of inliers. These methods are robust and accurate enough to guarantee the stability, but the computational efficiency is intractable for the time critical applications, *e.g.*, SLAM and navigation.

A remarkable fact is that the whole rotation space is quite redundant for all MF rotations, which is one of the main factors that disturbs the computational efficiency of the parameter search-based methods (typically using a branch-and-bound (BnB) framework [14]). The goal of this paper is to delimit a special sub-space that just contains all MF rotations (hereinafter called the "MFR space"). To achieve this, we prove that the whole rotation space can be evenly divided into 24 congruent MFR spaces, any of which just contains all MF rotations. Within a delimited MFR space, the MF estimation approach, *e.g.*, BnB-based algorithm, guarantees globally optimal solution satisfying both stability and efficiency. Moreover, based on the theory of rotation space delimitation, the general rotation estimation problems solved by the BnB-based algorithm associated with a rotation space search, *e.g.*, the panorama stitching [15, 17, 35], the estimation of camera pose [13, 26], the 3D registration [15, 33, 34] and the tilt-pan camera calibration [4, 5, 6], can be improved greatly. Overall, the main contributions of this paper are summarized as follows:

- We prove the redundancy of search in whole rotation space for the MF estimation. The volume of the MFR space is only 1 / 24 of that of the whole rotation space

theoretically.

- We prove that a cube with the half side length $\pi/4$ in the rotation space parametrized by angle-axis representation is just enough to tightly enclose a MFR space, and then an exact MFR space is delimited.

- Searching in the delimited MFR space, the efficiency of MF estimation based on BnB is improved by about **25** times, which satisfies both stability and efficiency of this problem. Furthermore, the general rotation problems associated with a whole rotation space search are solved more efficiently.

2. Related Work and Prerequisite Knowledge

Since many existing MF estimation methods search in parameter space to meet stability, we first review the related works, and then, the parameterization of rotation space.

2.1. Existing methods

The expectation maximization (EM)-based methods [18, 19, 20], cluster image lines and estimate VPs alternately. It alternates between an expectation (E) step, which estimates the line clustering given the current hypothesized VPs, and a maximization (M) step, which computes the VPs given the line clustering estimated at the E step. These methods are sensitive to the initial solution and prone to converging to a local optimum. The sampling-based methods [3, 21, 22, 23] exploit RANSAC [24] or its variants [25]. Bazin and Pollefeys [21] proposed a 3-line based algorithm for calibrated cameras, and based on this strategy, Zhang *et al.* [23] solved the 3-line hypothesis in a better approach that achieves a better level of confidence with fewer samples. Note that these methods are inherently unstable and cannot guarantee the global optimality.

To guarantee stability, Bazin *et al.* [15, 16] proposed orthogonal VPs estimation methods based on a BnB framework, which satisfies the globally optimality in terms of maximizing the number of inliers. In [16], the rotation space is parametrized by Euler angles representation, but this method based on interval analysis provides loose bounds. Concurrently, an improved strategy presented in [15] by exhaustively searching in the whole rotation space parametrized by angle-axis representation [26]. However, these BnB frameworks are usually too slow for time critical applications. Joo *et al.* [27, 28] recently proposed a novel strategy to significantly improve the efficiency of BnB. It transfers input data to an extended Gaussian image domain and combines with efficient bound functions to satisfy both stability and efficiency. However it is inherently suitable for surface normals, but not well applicable to image lines.

To meet the two properties, Bazin *et al.* [29] proposed to sample numerous MF rotations over the rotation space
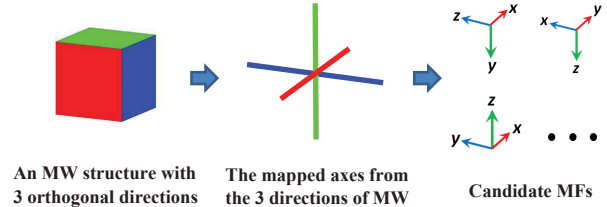

Figure 1. An MW structure maps to 24 candidate MFs with no prior information regarding the axis direction of the MF and the permutation of three axes.

around an initial rotation (*i.e.*, the quasi-exhaustive search) and select the one maximizing the number of inliers. While this tracking based strategy is inherently suitable for video sequences. More recently, Li *et al.* [30, 31] proposed a quasi-globally optimal algorithm by sampling one or two degrees of freedom (DOF) of MF rotation hypothesis and then searching for the other DOF with the BnB-based strategy. The method reaches a compromise between efficiency and accuracy by hybridizing the sampling and search strategies ingeniously, however, it can not completely satisfy the optimality because of the sampling mechanism.

2.2. Parameterization of Rotation Space

The rotation space parametrized by unit quaternion representation can be denoted as $\mathcal{H}$, which is a hemisphere in 4-space. It is isometric between $\mathcal{H}$ and $SO(3)$ [26]. Consider unit quaternions q_1 and q_2, and their corresponding rotations R_1 and R_2 respectively. The two metrics are related as follows:

$$d_R(R_1, R_2) = 2d(q_1, q_2) = 2\arccos(q_1 \cdot q_2) \quad (1)$$

where $d_R(R_1, R_2)$ is the rotation angle $\angle(R_1, R_2)$ between the two matrices, and symbol "·" represents the dot product.

The rotation space parametrized by angle-axis representation can be denoted as B_π, which is a ball of radius π. When mapping from $\mathcal{H}$ to B_π, it causes tangential stretching at the periphery [26]. It is not isometric between B_π and $SO(3)$. To ensure isometry, we still measure the distance between two the vectors r_1 and r_2 in B_π by the distance between the corresponding unit quaternions q_1 and q_2:

$$d_r(r_1, r_2) = 2d(q_1, q_2) = 2\arccos(q_1 \cdot q_2) \quad (2)$$

Where $d_r(r_1, r_2)$ is distance between the vectors r_1 and r_2. For the sake of unity of form, we denote $d(\cdot)$ as the distance metric on rotations regardless of the parameterization.

3. Problem Statement

Let's first state the redundancy of the whole rotation space for solving MF rotation, and then, prove that the rotation space can be divided into 24 congruent MFR spaces.

3.1. Redundant Search Space

With no prior information regarding the axis direction of the MF, a direction of an MW structure can be mapped to one of two opposite axes arbitrarily and we finally obtain 6 mapped candidate axes, as shown in Fig. 1. On the constraints of the right-hand rule and without prior information regarding the permutation of three axes of the MF, any combination of three orthogonal candidate axes can be regarded as a candidate MF. It is provable there are 24 such candidate MFs at all, and any candidate MF can interpret this MW structure. Therefore mapping from MW structure to MF is a 1-to-24 mapping.

Suppose the camera doesn't rotate with respect to the MW structure, *i.e.*, the camera is perspective and the projections of the parallel structural lines of the MW structure remain parallel in the image plane. In this case, the three directions of the MW structure are considered to align along or against the three axes of camera frame and the 6 mapped candidate axes corresponding to the three directions can be expressed as:

$$E = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \qquad (3)$$

where each column represents an axis of the MF. Obviously, on the constraints of the right-hand rule, different combinations of any three orthogonal axes of E can form 24 candidate MFs. We define these MFs as reference frames (RFs). By regarding these RFs as rotation matrices, we obtain a set $S = \{T_i\}_{i=0}^{23}$, where $T_i \in SO(3)$ is a RF and the RF T_0, which consists of a set $\{e_i\}_{i=1}^{3}$ [1] of three canonical vectors, coincides with the camera frame.

When the orientation of the camera is rotated δR (*i.e.*, called "MF rotation"), these six mapped axes of E will appear rotated by δR in the camera coordinate system:

$$M = \delta R E \qquad (4)$$

With the same combination rules as mentioned above, different combinations of any three orthogonal axes of M form a set of 24 unknown-but-sought candidate MFs $\{R_i\}_{i=0}^{23}$, where $R_i \in SO(3)$ is a candidate MF corresponding to the RF T_i. As a result, the MF rotation δR can be regarded as occurring between any unknown-but-sought candidate MF R_i and its corresponding RF T_i simultaneously, *i.e.*, the mapping from the MF rotation δR to the candidate MFs $\{R_i\}_{i=0}^{23}$ is a 1-to-24 mapping:

$$\delta R = R_i T_i^{-1}, \quad T_i \in S \qquad (5)$$

Then, we get the relationship between the candidate MF R_0 and any other candidate MF R_i:

$$R_i = R_0 T_0^{-1} T_i = R_0 T_i, \quad i \neq 0 \qquad (6)$$

---

[1] *i.e.*, $e_1 = [1\,0\,0]^T$, $e_2 = [0\,1\,0]^T$ and $e_3 = [0\,0\,1]^T$

Obviously, when evaluate the MF R_i, typically in terms of maximizing the number of inliers, we can obtain the same value. In theory, the whole rotation space search strategy based on the BnB framework [15] converges to 24 different positions concurrently. Therefore, The whole rotation space is quite redundant for solving the MF rotation.

3.2. Congruent MFR Spaces in Rotation Space

According to Sec. 2.2, the distance metric on rotations is in the form of unit quaternion representation uniformly. Given an MF R, we can always find a nearest RF such that:

$$d_{min} = \min_{T_i \in S} d(R, T_i) \qquad (7)$$

We call the MF R falls in the radiation range of the nearest RF T_{min}, and we denote the rotation between R and its nearest RF T_{min} as δR_{min}:

$$\delta R_{min} = R T_{min}^{-1}, \quad T_{min} \in S \qquad (8)$$

On the contrary, given a RF T_i, there is always a set of corresponding MFs, which fall in the radiation range of T_i.

Result 1. The whole rotation space D can be divided into 24 sub-spaces $\{D_i\}_{i=0}^{23}$ (*i.e.*, called "MFR spaces"), *i.e.*, $D = \bigcup_{i=0}^{23} D_i$, where the MFR space D_i is centered at the RF T_i.

According to Eqs. (5) and (8), given an MF R in the MFR space D_{min} centered at the RF T_{min}, we can always find a candidate MF R_i in any other MFR space D_i centered at the RF T_i such that the RF T_i is rotated to R_i by the same rotation δR_{min}:

$$\delta R_{min} = R T_{min}^{-1} = R_i T_i^{-1}, \quad i \neq min \qquad (9)$$

Result 2. The volume of these 24 MFR spaces is congruent and each of these MFR spaces just contains all MF rotations, *i.e.*, $V(D_i) = V(D_j), \forall i, j = 0...23 \;\; and \;\; i \neq j$, where $V(\cdot)$ is the volume of MFR space. Note that those rotations, which fall in more than one MFR space, distribute at the common boundaries of these sub-spaces. However they can be imposed to attribute to one of the corresponding MFR spaces uniquely and evenly so as to there is no overlap between these MFR spaces, *i.e.*, $D_i \bigcap D_j = \varnothing, \quad i \neq j$.

4. Delimitation of MFR Space

This section states how to delimit an MFR space. First, we define a cube of just the right size to enclose one of the MFR spaces. Secondly, we delimit an exact MFR space.

4.1. A Cube Enclosing an MFR Space

For the sake of visual expression, we employ the angle-axis representation to parameterize the rotation space. We denote the RFs in the set S as 24 anchors fixed in the ball
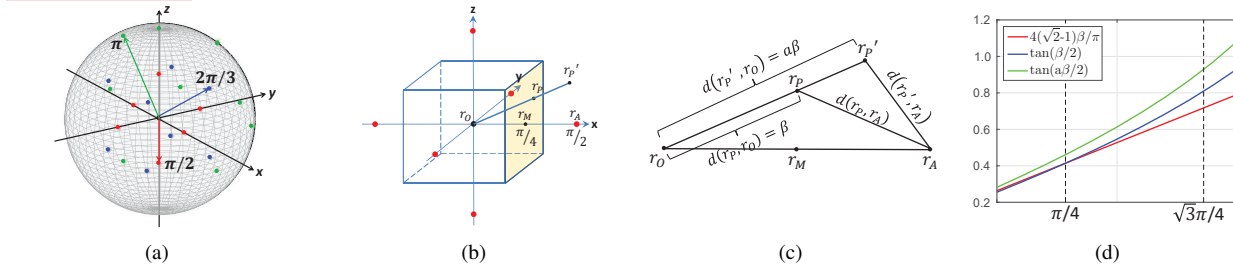
Figure 2. Determination of the cube C_O enclosing the MFR space D_O. (a) Distribution of the anchors in the rotation space B_π. (b) The cube C_O enclosing the MFR space D_O centered at r_O. (c) Comparison of the distances from points on and outside the plane $x = \pi/4$ to the point r_O and the point r_A. (d) Comparison of equivalent function values of the corresponding distances.

B_π, as shown in Fig. 2(a). Then the angle-axis parameterized form of T_0 can be denoted by the anchor r_O, which is at the coordinate origin O. In addition, 6 anchors (red points), $\pi/2$ away from it, are distributed on three axes, 8 anchors (blue points), $2\pi/3$ away from it, are distributed in 8 quadrants and 9 anchors (green points), π away from it, are distributed on the surface of the ball B_π.

Since the 6 anchors on axes are nearest to the anchor r_O, we select one of them to analyze, *e.g.*, the anchor r_A at coordinate $(\pi/2, 0, 0)$ is selected. Obviously, the point r_M, which is at the midpoint of the line between r_O and r_A, is equidistant to r_O and r_A (*i.e.*, the distances $d(r_M, r_O)$ and $d(r_M, r_A)$ are both $\pi/4$). According to **Result 2**, the point r_M is on the common boundary of the MFR space D_O centered at r_O and the MFR space D_A centered at r_A. It is provable that the points on the line between r_O and r_M are nearest to the anchor r_O. Therefore, when use a cube C_O to enclose the the MFR space D_O, *the half side length of the cube C_O cannot be less than $\pi/4$*, see Fig. 2(b). But is the cube C_O big enough to enclose the MFR space D_O?

According to the nearest distance criterion, if the distance from a point r_P on a plane of the cube C_O to the anchor r_O is greater than that from r_P to some other anchors, we can conclude that the point r_P is out of the MFR space D_O. Then, if all the points on and outside all the 6 planes of the cube C_O are out of the MFR space D_O, the closed space enclosed by the cube C_O is big enough to hold the MFR space D_O.

i) Points on the planes of the cube. For the sake of simplicity, we focus on the plane $x = \pi/4$ passing through the point r_M, and a point r_P on the plane can be expressed as $r_P = (\pi/4, y, z)$, where $y \in [-\pi/4, \pi/4]$ and $z \in [-\pi/4, \pi/4]$. The quaternion form of the point r_P is $q_P = (\cos(\beta/2), \frac{\pi \sin(\beta/2)}{4\beta}, \frac{y \sin(\beta/2)}{\beta}, \frac{z \sin(\beta/2)}{\beta})$, where $\beta = \sqrt{(\pi/4)^2 + y^2 + z^2}$, $\beta \in [\pi/4, \sqrt{3}\pi/4]$. The distance from the point r_P to anchor r_O is:

$$d(r_P, r_O) = \beta \tag{10}$$

and the distance from the point r_P to the anchor r_A is:

$$d(r_P, r_A) = 2\arccos\left(\frac{\sqrt{2}\cos(\beta/2)}{2} + \frac{\sqrt{2}\pi \sin(\beta/2)}{8\beta}\right) \tag{11}$$

Our goal is to compare $d(r_P, r_O)$ and $d(r_P, r_A)$, see Fig. 2(c). After some equivalent transformations, comparing Eq. (10) and Eq. (11) is equivalent to comparing the values of the two functions $\tan(\beta/2)$ and $4(\sqrt{2} - 1)\beta/\pi$. Fig. 2(d) shows that in the domain of parameter β, the value of function $\tan(\beta/2)$ is greater than that of function $4(\sqrt{2}-1)\beta/\pi$, except for the value of $\beta = \pi/4$, where $d(r_P, r_O)$ is equal to $d(r_P, r_A)$, *i.e.*, the point r_P is at the point r_M.

ii) Points outside the planes of the cube. The points outside the planes of the cube can be considered to be on the extension line from the anchor r_O to the corresponding points on the plane, see Fig. 2(b). Consider a point r_P' on the extension line from r_O to r_P, which is expressed as $r_P' = (a\pi/4, ay, az)$, where $a > 1$, and the quaternion form of the point r_P' is $q_P = (\cos(a\beta/2), \frac{\pi \sin(a\beta/2)}{4\beta}, \frac{y \sin(a\beta/2)}{\beta}, \frac{z \sin(a\beta/2)}{\beta})$. The distance from the point r_P' to anchor r_O is:

$$d(r_P', r_O) = a\beta \tag{12}$$

and the distance from the point r_P' to the anchor r_A is:

$$d(r_P', r_A) = 2\arccos\left(\frac{\sqrt{2}\cos(a\beta/2)}{2} + \frac{\sqrt{2}\pi \sin(a\beta/2)}{8\beta}\right) \tag{13}$$

In the same way as above, comparing Eq. (12) and Eq. (13) is equivalent to comparing the values of the two functions $\tan(a\beta/2)$ and $4(\sqrt{2} - 1)\beta/\pi$, and the value of function $\tan(a\beta/2)$ is greater than that of function $4(\sqrt{2} - 1)\beta/\pi$, see Fig. 2(c,d). With these analyses, we conclude that the plane $x = \pi/4$ is externally tangent to the MFR space D_O.

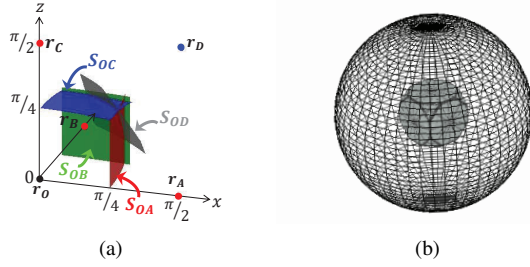**Result 3**. The cube C_O with the half side length $\pi/4$ is just enough to tightly enclose the MFR space D_O.

Figure 3. Definition procedure of the MFR space D_O. (a) The common boundaries of the MFR space D_O and its adjacent MFR spaces in the 1st quadrant. (b) The outer boundary of the MFR space D_O centered at the anchor r_O in the rotation space B_π.

4.2. An Exact MFR Space

Consider the 8 anchors in the quadrants, *e.g.*, the anchor r_D in the 1st quadrant. The midpoint of the line between r_O and r_D are equidistant to r_O and r_D. Since this midpoint is in the cube C_O, the MFR space D_O intersects with the MFR space D_D centered at r_D, see Fig. 3(a). While it does not intersect with the MFR spaces centered at the anchors distributed on the surface of the ball B_π. So we just consider the MFR space D_O and the MFR spaces centered at the anchors distributed on the axes and in the quadrants.

We choose three anchors r_A, r_B and r_C on positive of axes and the anchor r_D in the 1st quadrant to illustrate. Let's focus on the MFR space D_O and one of its 4 neighboring MFR spaces, *e.g.*, the MFR space D_A centered at the anchor r_A. The point r_i on the common boundary between the two MFR spaces D_O and D_A satisfies:

$$d(r_O, r_i) = d(r_A, r_i) \qquad (14)$$

This common boundary forms a smooth surface S_{OA}, and similarly, the other common boundaries between the MFR space D_O and the other three neighboring MFR spaces D_B, D_C and D_D form three surfaces S_{OB}, S_{OC} and S_{OD} respectively. Finally, the four surfaces delimit the sub-space belonging to D_O in the 1st quadrant, as shown in Fig. 3(a).

Based on the symmetry, the MFR space D_O intersects with its 14 surrounding MFR spaces to form 14 common boundaries, and the closed space enclosed by them is the exact MFR space D_O, as shown in Fig. 3(b). Therefore, we can judge whether a point in the cube C_O is within the MFR space D_O by simply comparing its distance to r_O and these surrounding anchors.

5. Efficient and Stable MF Estimation

This section details the algorithm procedure of MF estimation based on the proposed method, and then extend it to solve the general rotation problems.

Algorithm 1 The efficient and stable MF estimation.
Initialize the cube list $\mathcal{L}$ with C_O $s.t.$ $C_O = (\pi/2)^3$.
repeat
  Subdivide each cube of $\mathcal{L}$ congruently ($\sigma \leftarrow \sigma/2$).
  for each cube C_i in $\mathcal{L}$ **do**
    if $L^* > P(C_i)$ **or** C_i outside the MFR space D_O **do**
      continue
    end if
    Calculate the rotation $\overline{R_i}$ at the center of the cube C_i.
    Compute lower bounds $L(\overline{R_i})$ and upper bounds $U(\overline{R_i})$.
    Update L^* to $L^* = \max(L^*, L(\overline{R_i}))$.
  end for
  $i^* = \arg\max_i U(\overline{R_i})$, $U^* = U(\overline{R_{i^*}})$, $R^* = \overline{R_{i^*}}$.
  Remove all the cubes from $\mathcal{L}$ such that $U(\overline{R_i}) < L^*$.
until $\exists i$, such that $L(\overline{R_i}) = U^*$ or it reaches a desired accuracy.
Return: R^* (*i.e.*, the rotation maximizing the number of inliers).

5.1. Algorithm Procedure

For the MF estimation, Bazin *et al.* proposed a BnB-based method [15] to guarantee globally optimal solution with high stability. The proposed method improves the solution of the problem to satisfy both stability and efficiency.

Given a set of image lines, we aim to find which line belongs to which axis of the unknown-but-sought MF R. On the Gaussian sphere, a image line l_i is back-projected onto the sphere as a great circle, which is represented by a normal vector n_i. Let $v_i = Re_i$ be one axis and e_i belongs to a set $\{e_i\}_{i=1}^3$ of three canonical vectors. Concretely, we consider that the normal-axis pair (n_i, Re_j) is an inlier if their geometric distance is lower than a residual tolerance τ, *i.e.*, $|\angle(n_i, Re_j) - \pi/2| < \tau$. The consensus set maximization can be written:

$$
\begin{aligned}
\max_{\{y_{ij}\}, R} \quad & \sum_{i=1}^{N} \sum_{j=1}^{3} y_{ij} \\
s.t. \quad & y_{ij}|\angle(n_i, Re_j) - \pi/2| < y_{ij}\tau, \\
& y_{ij} \in \{0, 1\}, \forall i = 1 \ldots N, j = 1, 2, 3
\end{aligned}
\qquad (15)
$$

where y_{ij} is auxiliary variable introduced for indicating whether the normal-axis pair (n_i, Re_j) is an inlier ($y_{ij} = 1$) or an outlier ($y_{ij} = 0$). To solve Eq. (15), the basic idea of BnB is to divide the MFR space into smaller sub-spaces and remove the spaces that cannot contain a solution better than the current one. This removing decision is made by a feasibility test and the associated bounds. Iteratively, the size of the sub-spaces decreases and the estimated solution converges to the optimal solution.

In branching part, let C_O be an initial cube that tightly encloses the MFR space D_O, and then divide the search space into smaller congruent sub-spaces by octal subdivision of the cube. In bounding part, we re-state the bound computation suggested by Bazin *et al.* [15]. Given a cube

C with the half side length σ, we get the MF $\overline{R}$ corresponding to the center of it. The lower bound $L(\overline{R})$ of the cube can be written:

$$\max_{\{y_{ij}\}} \quad \sum_{i=1}^{N}\sum_{j=1}^{3} y_{ij}$$
$$s.t. \quad y_{ij}|\angle(n_i, \overline{R}e_j) - \pi/2| < y_{ij}\tau, \quad (16)$$
$$y_{ij} \in \{0,1\}, \forall i = 1\dots N, j = 1,2,3$$

The upper bound $U(\overline{R})$ can be written:

$$\max_{\{y_{ij}\}} \quad \sum_{i=1}^{N}\sum_{j=1}^{3} y_{ij}$$
$$s.t. \quad y_{ij}|\angle(n_i, \overline{R}e_j) - \pi/2| < y_{ij}(\tau + \sqrt{3}\sigma), \quad (17)$$
$$y_{ij} \in \{0,1\}, \forall i = 1\dots N, j = 1,2,3$$

The solutions of Eqs. (16) and (17), *i.e.*, $L(\overline{R})$ and $U(\overline{R})$, are simply obtained by exhaustively checking the inlier constraint for each normal vector with respect to a given rotation $\overline{R}$ that corresponds to the center of a given cube C

A typical breadth-first-search (BFS) tests the feasibility of each cube in the list $\mathcal{L}$ layer by layer iteratively. A more efficient strategy is that each sub-cube C_i inherits its parent's upper bound $P(C_i)$. At the beginning of each iteration, the maximum lower bound L^* is first updated by the cubes with the maximum upper bound and/or the maximum lower bound, and subsequently, with the real-time update of L^*, each other cube uses its inherited upper bound to test its feasibility firstly, if feasible, compute its own bounds. The improved BnB procedure is formalized in Alg. 1.

5.2. Extension to Rotation Estimation

Let u_i and v_i be the ith matching pair of two sets of 3D points related by a rotation R and we define the matching pair (u_i, v_i) an inlier if the distance is lower than a residual tolerance τ, *i.e.*, $\angle(v_i, Ru_i) < \tau$. Then we modify the formulation of Eq. (15) to general rotation estimation:

$$\max_{\{y_i\},R} \quad \sum_{i=1}^{N} y_i$$
$$s.t. \quad y_i\angle(v_i, Ru_i) < y_i\tau, \quad (18)$$
$$y_i \in \{0,1\}, \forall i = 1\dots N$$

To solve Eq. (18), according to **Result 1** and **Result 2**, the BnB-based algorithm associated with a whole rotation space search can be modified to search in the 24 MFR spaces theoretically.

Given a rotation R_0 in the MFR space D_O, it can be moved to any other candidate MF rotation R_i in the corresponding MFR space D_i according to Eq. (6). Then the
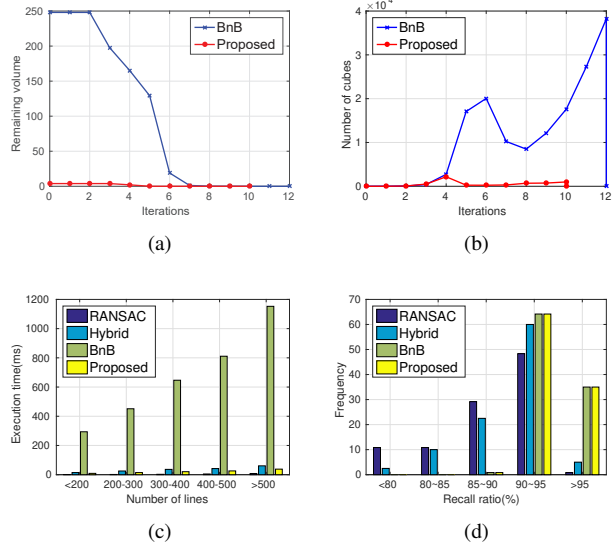

(a)

(b)

(c)

(d)

Figure 4. Efficiency and accuracy comparison on synthesized data for MF estimation. (a) Evolution of the volume of the search space. (b) Evolution of the number of tested cubes. (c) Execution time of various methods w.r.t. the number of image lines. (d) The distribution of recall ratio of various methods.

Method	θ_x	θ_y	θ_z	θ
RANSAC [23]	$2.00°$	$1.65°$	$1.39°$	$2.19°$
Hybrid [30]	$1.07°$	$0.93°$	$1.01°$	$1.28°$
BnB [15]	$0.61°$	$0.56°$	$0.47°$	$0.71°$
Proposed	$0.60°$	$0.54°$	$0.44°$	$0.67°$

Table 1. Comparison of the mean angular errors on synthesized data for MF estimation.

MFR space D_O can be moved to the location of the corresponding MFR space D_i by this transformation. Finally, we complete the exhaustive search of the whole rotation space by only searching in D_O.

Since tangential stretching at the periphery of the rotation space parametrized by angle-axis representation [26], when use an initial cube with the half side length π to enclose the ball B_π [15], the search space of the exhaustive BnB framework is further expanded to $8\pi^3$. However, when use the cube C_O to enclose the MFR space D_O to complete the whole rotation space search, the total volume of the search space is equivalent to $24 \times (\pi/2)^3$, which is far less than $8\pi^3$. Therefore, the efficiency of the general rotation estimation will be improved theoretically. Experimental results illustrate the performance of the improved procedure.

6. Experiments

We present our experimental results to explore the performance of the proposed method and to compare with the state-of-the-art methods:

Figure 5. Some example images obtained by the exhaustive BnB and the proposed method on YUD [18]. The first row shows the example images with extracted lines. The second row shows the line classification results of the exhaustive BnB. The third row shows the line classification results of the proposed method.

- For MF estimation, besides the exhaustive **BnB** [15], we also compare two most recently reported algorithms featuring remarkable performance: **RANSAC** [23] that retrieves the optimal MF rotation hypothesized by 3 sampled image lines and **Hybrid** [30] that samples two DOF of MF rotation hypothesis and searches for the optimal third DOF based on BnB.

- For the general rotation estimation, we show the performance of the proposed method by comparing with the exhaustive **BnB** [15].

All the mentioned methods are implemented in C++ and run on a laptop equipped with an Intel i5-7360U 2.3GHz CPU (a single core was used) and 8GB RAM.

Given a ground truth MF rotation matrix and its estimated MF rotation matrix, we evaluate the MF estimation accuracy in terms of 1) the recall ratio [32] defined by $N_c/(N_c + N_m)$, where N_c and N_m are the numbers of correctly clustered inliers and missing inliers respectively; 2) the mean angular error θ of the rotation between the two matrices; 3) the mean angular errors θ_x, θ_y and θ_z between the axes of the estimated MF and that of the ground truth. Similarly, we use criteria 1) and 2) to evaluate the accuracy of the general rotation estimation.

6.1. Synthesized Data: MF Estimation

The number of iterations of **RANSAC** is adaptively selected from 40 to 100 and the number of samples of **Hybrid** is automatically adjusted according to the outliers ratio, which is set to 30%. For the exhaustive **BnB** and the
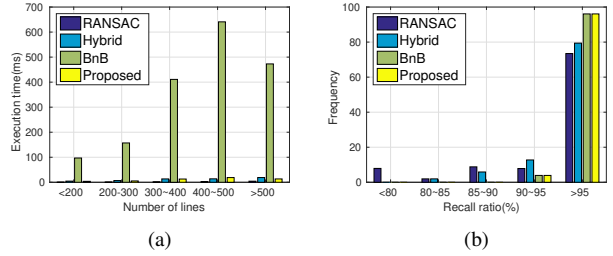

Figure 6. Efficiency and accuracy comparison on YUD [18]. (a) Execution time of various methods w.r.t. the number of image lines. (b) The distribution of recall ratio of various methods.

proposed method, the desired accuracy (the side length of cube) is set to $0.1°$. Accordingly, we define the residual tolerance $\tau = 2°$. In the synthetic simulation, we perform various experiments to demonstrate accuracy, convergence and efficiency of the proposed method. Three mutually orthogonal unit vectors corresponding to three axes of MF are synthesized, and three sets of orthogonal parallel 3D lines are generated according to these axes. Then three sets of image lines are obtained by the projection of a synthetic camera. We perturb the endpoints of these image lines by a zero-mean Gaussian noise with standard deviation of 3 pixels, and corrupt the outlier ratio 30% to generate outlier image lines by randomizing their positions and directions.

Fig. 4(a,b) shows a representative comparison with 200 inlier lines. The volume of initial search space of the proposed method is only 1 / 64 of that of the exhaustive **BnB** (Fig. 4(a)), and the number of tested cubes of the proposed method is far less than that of the exhaustive **BnB** (Fig. 4(b)). Therefore, the proposed method maintains a relatively lower level for memory utilization. For statistical comparison, with the same outlier ratio and noise level, we generate 100 synthesized data whose inlier lines ranges from tens to hundreds. We evaluate the execution time with respect to the number of inlier lines (Fig. 4(c)) and the recall ratio (Fig. 4(d)). The comparison of the several angluar errors of these methods is listed in the Table 1. These experimental data show that the proposed method reveals efficient computation time and reliable accuracy.

6.2. Real Data: MF Estimation

We test the proposed method on the York Urban Database (YUD) [18] which is composed of 102 images acquired in man-made environments. Each image contains a set of manually extracted lines corresponding to 2 or 3 VPs. For each line clustered by the proposed method, we successfully obtain the true clustering, which demonstrates the correctness of our strategy. Fig. 5 shows some representative results obtained by the exhaustive **BnB** and the proposed method. Fig. 6(a) illustrates the distribution of the
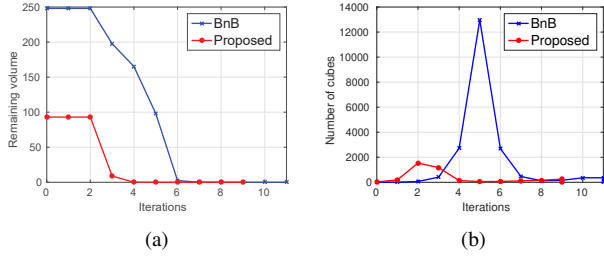
(a) (b)

Figure 7. Remaining volume and tested cubes comparison on synthesized data for rotation estimation. (a) Evolution of the volume of the search space. (b) Evolution of the number of tested cubes.

Method	θ		Recall ratio		Time
	Mean	Median	Mean	Median	
BnB[15]	$0.63°$	$0.62°$	99.11%	99.00%	40.43ms
Proposed	$0.59°$	$0.58°$	99.10%	99.00%	9.29ms

Table 2. Comparison of the mean angular error θ, the recall ratio and runtime on synthesized data for rotation estimation.

execution time, and Fig. 6(b) reports the recall ratio of various methods on all the images. The exhaustive **BnB** and the proposed method stably obtain all the inliers, however the running time of the proposed method is about 25 times faster than that of the exhaustive **BnB** statistically.

6.3. Synthesized Data: Rotation Estimation

In the coordinate interval of $[-2, 2] \times [-2, 2] \times [-2, 2]$, we randomly generate two sets of 300 points related by any random 3D rotation. We apply a zero-mean Gaussian noise, whose standard deviation is 0.1, to the coordinates of these points and corrupt 20% of these points to create outliers. For the comparison methods, the desired accuracy is $0.1°$, and the residual tolerance τ is $2°$.

The volume of initial search space of the proposed method is far less than that of the exhaustive **BnB**. Fig. 7(a,b) illustrates the evolution of the volume of the search space and the number of tested cubes. It shows the proposed method converges to the optimal solution with less computation. For statistical comparison, with the same outlier ratio and noise level, we generate 100 synthesized data whose inlier point-point pair ranges from tens to hundreds. Table 2 shows the performance comparison of the two methods. The efficiency of the proposed method is about 4 times faster than that of the exhaustive **BnB** statistically.

6.4. Real Data: Panorama Stitching

We use the PASSTA dataset to test the proposed method in panorama stitching applications [35]. To obtain correspondences between the two images, we extract and match SIFT features [36] and apply the inverse of the known camera intrinsic matrix K to obtain unit-norm bearing vectors.
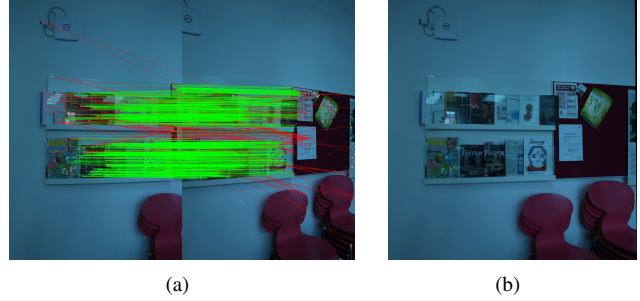

(a) (b)

Figure 8. Result of panoramic image stitching on LRBD [35]. (a) Inlier (green lines) and outlier (red lines) matches detected by the proposed method between the two input images. (b) Stitching of the two images to build a panoramic view.

Method	Recall ratio	Cubes	Time
BnB[15]	67.24%	5154	54.88ms
Proposed	67.24%	2595	26.77ms

Table 3. Comparison of the recall ratio, the average of the maximum number of tested cubes and runtime on LRBD [35]

Using the estimated rotation, we compute the homography matrix as $H = KRK^{-1}$ to stitch the pair of images together. For statistical comparison, we randomly select 30 pairs of images that have overlapping area. Fig. 8(a) illustrates the established correspondences between two input images from the Lunch Room Blue dataset (LRBD). The final stitching result is shown in Fig. 8(b). Table 3 shows the performance comparison of the two methods. We measure the memory usage by the average of the tested cubes, *i.e.*, the comparison item "Cubes" in Table 3. The proposed method performs same accurate stitching with less time.

7. Conclusions

Motivated by meeting both the stability and efficiency of MF estimation, we proposed a pioneering theory of rotation space delimitation that the whole rotation space can be evenly divided into 24 congruent MFR spaces. Based on this theory, the MF estimation method with BnB framework meets these two requirements. The broader impact of our theory is beneficial to improve the efficiency of general rotation estimation problems, *e.g.*, camera pose estimation, online camera calibration, 3D registration and panoramic stitching. We have demonstrated the validity of our theory by experimenting on various synthetic and real datasets.

Since our contribution is the delimitation of the MFR space, we search for rotation solution based on a simple and general BFS. Alternative search methods like Best-First-Search or A*, may speed up the solution of optimal rotation more or less and we leave this in the future.

References

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

[2] Horst Wildenauer and Allan Hanbury. Robust camera self-calibration from monocular images of Manhattan worlds. In *CVPR*, 2012. 1

[3] Jean-Philippe Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *ICCV*, 2009. 1, 2

[4] Yongduek Seo, Young-Ju Choi, and Sang Wook Lee. A branch-and-bound algorithm for globally optimal calibration of a camera-and-rotation-sensor system. In *ICCV*, 2009. 1

[5] Jan Heller, Michal Havlena, and Tomas Pajdla. A branch-and-bound algorithm for globally optimal hand-eye calibration. In *CVPR*, 2012. 1

[6] Jean-Charles Bazin, Yongduek Seo, Richard Hartley, and Marc Pollefeys. Globally optimal inlier set maximization with unknown rotation and focal length. In *ECCV*, 2014. 1

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

[8] A. Criminisi, I. Reid, and A. Zisserman . Single view metrology. *IJCV*, 40(2):123-148, 2000. 1

[9] Wongun Choi, Yu-Wei Chao, Caroline Pantofaru, and Silvio Savarese. Understanding indoor scenes using 3D geometric phrases. In *CVPR*, 2013. 1

[10] 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

[11] Yanyan Li, Nikolas Brasch, Yida Wang, Nassir Navab, and Federico Tombari. Structure-SLAM: Low-drift monocular SLAM in indoor environments. *RAL*, 2020. 1

[12] Huizhong Zhou, Danping Zou, Ling Pei, Rendong Ying, Peilin Liu, and Wenxian Yu. StructSLAM: Visual SLAM with building structure lines. *IEEE Transactions on Vehicular Technology*, 64(4):1364-1375, 2015. 1

[13] Dylan Campbell, Lars Petersson, Laurent Kneip, and Hongdong Li. Globally-Optimal Inlier Set Maximisation for Simultaneous Camera Pose and Feature Correspondence. In *ICCV*, 2017. 1

[14] R. Horst and H. Tuy. Global optimization: Deterministic approaches. *Springer Science & Business Media*, 2013. 1

[15] Jean-Charles Bazin, Yongduek Seo, and Marc Pollefeys. Globally optimal consensus set maximization through rotation search. In *ACCV*, 2012. 1, 2, 3, 5, 6, 7, 8

[16] 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

[17] Matthew Brown and David G. Lowe. Automatic panoramic image stitching using invariant features. *IJCV*, 74(1):5973, 2007. 1

[18] Patrick Denis, James Elder, and Francisco Estrada. Efficient edge-based methods for estimating Manhattan frames in urban imagery. In *ECCV*, 2008. 2, 7

[19] Julian Straub, Guy Rosman, Oren Freifeld, John J. Leonard, and John W. Fisher. A mixture of Manhattan frames: Beyond the Manhattan world. In *CVPR*, 2014. 2

[20] Julian Straub, Oren Freifeld, Guy Rosman, John J. Leonard, and John W. Fisher. The Manhattan frame model-Manhattan world inference in the space of surface normals. *TPAMI*, 99:235-249, 2018. 1, 2

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

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

[23] 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*, 117(2):111-130, 2016. 2, 6, 7

[24] 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*, 24(6):381395, 1981. 2

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

[26] Richard I. Hartley and Fredrik Kahl. Global optimization through rotation space search. *IJCV*, 82(1):6479, 2009. 1, 2, 6

[27] Kyungdon Joo, Tae-Hyun Oh, Junsik Kim, and In So Kweon. Globally optimal manhattan frame estimation in real-time. In *CVPR*, 2016. 2

[28] Kyungdon Joo, Tae-Hyun Oh, Junsik Kim, and In So Kweon. Robust and globally optimal Manhattan frame estimation in near real time. *TPAMI*, 41(3):682-696, 2019. 2

[29] Jean-Charles Bazin, Cedric Demonceaux, Pascal Vasseur, and Inso Kweon. Rotation estimation and vanishing point extraction by omnidirectional vision in urban environment. *IJRR*, 31(1):63-81, 2012. 2

[30] Haoang Li, Ji Zhao, Jean-Charles Bazin, Wen Chen, Zhe Liu, and Yun-Hui Liu. Quasi-globally optimal and efficient vanishing point estimation in Manhattan world. In *ICCV*, 2019. 2, 6, 7

[31] Haoang Li, Ji Zhao, Jean-Charles Bazin, and Yun-Hui Liu. Quasi-globally Optimal and Near/True Real-time Vanishing Point Estimation in Manhattan World. *TPAMI*, 38(11):22412254, 2020. 2

[32] C. M. Bishop Pattern Recognition and Machine Learning *Springer*, 2013. 7

[33] Jiaolong Yang, Hongdong Li, and Yunde Jia. Go-ICP: Solving 3D registration efficiently and globally optimally. In *ICCV*, 2013. 1

[34] Jiaolong Yang, Hongdong Li, Dylan Campbell, and Yunde Jia. Go-ICP: A globally optimal solution to 3D ICP point-set registration. *TPAMI*, 38(11):22412254, 2016. 1

[35] Giulia Meneghetti, Martin Danelljan, Michael Felsberg, and Klas Nordberg. Image alignment for panorama stitching in sparsely structured environments. *Scandinavian Conference on Image Analysis*, 2015. 1, 8

[36] David G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91-110, 2004. 8