Optimal Essential Matrix Estimation via Inlier-Set Maximization

Jiaolong Yang[1,2], Hongdong Li[2], Yunde Jia[1]

[1] Beijing Laboratory of Intelligent Information Technology,
Beijing Institute of Technology
[2] Australian National University and NICTA Australia
{yangjiaolong,jiayunde}@bit.edu.cn, hongdong.li@anu.edu.au

Abstract. In this paper, we extend the globally optimal "rotation space search" method [11] to essential matrix estimation in the presence of feature mismatches or outliers. The problem is formulated as inlier-set cardinality maximization, and solved via branch-and-bound global optimization which searches the entire essential manifold formed by all essential matrices. Our main contributions include an explicit, geometrically meaningful essential manifold parametrization using a 5D direct product space of a solid 2D disk and a solid 3D ball, as well as efficient closed-form bounding functions. Experiments on both synthetic data and real images have confirmed the efficacy of our method. The method is mostly suitable for applications where robustness and accuracy are paramount. It can also be used as a benchmark for method evaluation.

Keywords: Essential matrix, robust estimation, global optimization, branch-and-bound

1 Introduction

Essential matrix estimation is a basic building block for Structure from Motion (SfM). Given two views of a rigid scene from a calibrated perspective camera, the task is to estimate the relative pose or motion between the two views. Essential matrix can be estimated with image point correspondences using epipolar geometry. In reality, correspondence outliers are ubiquitous. For instance, natural or man-made scenes often contain similar structures, flat (and ambiguous) regions, repetitive patterns *etc.*, making flawless feature matching nearly impossible.

To deal with outliers in the context of multiple-view geometry, RANSAC [7] and its variants have played a major role. These methods, which are based on random sampling, cannot provide an optimality guarantee, and the inlier sets they find often vary from time to time. Moreover, in most RANSAC algorithms (*e.g.* [8,26]), to ensure efficiency an algebraic solver (*e.g.* the 5-point method [24,20]) and the 8-point method [21,10]) is often adopted to compute tentative estimation, followed by a thresholding stage using geometric reprojection error or Sampson error. The apparent inconsistency here, *i.e.* algebraic solver versus geometric threshold, can lead to inferior estimate.

In contrast, this work seeks a consistent, and globally optimal solution to essential matrix estimation, based on meaningful geometric error. By optimal, we adopt the consensus set maximization idea of RANSAC, *i.e.* to find the maximal-sized inlier set that is compatible with the input image measurements. To distinguish inliers from outliers, we use angular reprojection error. With a calibrated camera, it is natural to use angular reprojection error, because a calibrated camera behaves just like an angle measurement device, and every image point (represented by a unit vector) gives the actual viewing angle.

To achieve globally maximal inlier-set, a naive way would be exhaustively enumerating all possible combinations of inliers/outliers. However, this soon becomes intractable as combinations grow exponentially with point number. No efficient solver to this combinational problem exists to our knowledge. Our idea in this paper is: rather than searching over all *discrete* combinations of inliers, we search the entire *continuous* parameter space of essential matrix. To this end, it is necessary to find a suitable domain representation (parametrization) of the space, with which the bounds can be easily derived and efficiently evaluated.

The proposed method is based on systematically searching two (reduced) rotation spaces using branch-and-bound (BnB). It is inspired by the rotation search technique proposed in [11], which has been used in several vision problems [13,1,31]. To minimize the L_∞-norm of angular errors, [11] uses BnB to recursively search SO(3) with elegant bounding. However, L_∞-optimization is known to be extremely vulnerable to outliers, and [11] assumes outlier-free correspondences. Contrastly, our method works in the presence of outliers.

1.1   Related work

Our method is closely related to [11], and extends [11] to optimal inlier-set maximization which is non-trivial. A key insight for [11] to applying rotation search to essential matrix estimation is that, given rotation, the translation can be optimally solved with convex optimization (SOCP/LP). Contrastly, optimally solving the translation maximizing inlier-set cardinality is not trivial. The optimal essential matrix problem considered in this paper is more challenging. Method of [1] achieves inlier-set maximization with rotation search, however translation is assumed to be known. In this paper, we optimally solve the problem by searching the essential manifold with BnB, based on a novel parametrization scheme.

There have been some research efforts devoted to optimal essential matrix estimation with inlier-set maximization criterion [5,6]. Most closely related to our method is [5] in which a brute-force search method is proposed using triangulation feasibility test. The solution is exhaustively searched over the discretized parameter space formed by two unit spheres, and GPU implementation is used to speed up the computation. In [6], double pairs of correspondences are used, from which camera pose is found by searching the two epipoles via BnB. An approximation is made to solve an otherwise NP-hard problem (minimum vertex cover), which compromises the global optimality guarantee. The closed-form bounding functions we use in this paper are inspired by [5] (with necessary extension);

we however introduce other innovations in both parametrization scheme and optimization technique. By our method, an exact optimality can be achieved.

Some approaches use branch-and-bound methods for finding globally optimal fundamental matrix [19,32]. In particular, inlier-set is optimally maximized in [19] with algebraic error. Geometrically meaningful error is investigated in [32], but the goal is optimal error minimization assuming no outlier. These works discuss uncalibrated cases only, where the underlying Euclidean constraints of essential matrix are not exploited.

Another line of related work is outlier removal using convex optimization [28,16,18,25]. These methods are able to detect potential outliers with respect to a given threshold. However, the goal is not inlier-set maximization and outliers may be removed at the expense of losing some true inliers. Moreover, in SfM they assume known rotation to formulate the problem to be (quasi-)convex. Our work is also related to the study of SfM without pre-built correspondences [4,23], in a sense that we all compute the motion yielding most agreeable correspondences.

2   Essential Manifold Parametrization

A rigid motion comprises rotation and translation. As such, an essential matrix E relates to a 3D rotation $\hat{R} \in SO(3)$ and a 3D translation $\hat{\mathbf{t}} \in \mathbb{R}^3$ from the first camera to the second one by $E = [\hat{\mathbf{t}}]_\times \hat{R}$ where $[\cdot]_\times$ denotes the skew-symmetric matrix representation. Essential matrix can only be determined up to an unknown scale. To resolve this scale indeterminacy one can set the length of $\hat{\mathbf{t}}$, $i.e.$ $\|\hat{\mathbf{t}}\|$ to be fixed ($e.g.$ to be 1). Therefore, we have $\hat{\mathbf{t}} \in \mathbb{S}^2$, $i.e.$ a 2-sphere embedded in $\mathbb{R}^3$. In this way the essential manifold can be parameterized with 5 degrees of freedom (dofs) in $SO(3) \times \mathbb{S}^2$. In this paper we advocate different coordinate system and parametrization scheme to facilitate our BnB algorithm.

In solving the relative pose problem, one has the freedom to arbitrarily choose a coordinate system as the world frame. Different from a common practice which sets the first camera matrix to be $[\mathbf{I} \mid \mathbf{0}]$, we fix the first camera's center at the origin, $i.e.$ $\mathbf{C} \equiv \mathbf{0}$, and fix the second camera's center at $\mathbf{C}' \equiv [0,0,1]^T$ on the Z-axis.[1] We use R to denote the $absolute$ $orientation$ of the first camera (relative to the world frame), and R' for the second camera. Then, it is easy to see that, under this configuration the essential matrix can be written as

$$E = [-R'\mathbf{C}']_\times R'R^T = R'[-\mathbf{C}']_\times R^T = R' \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} R^T. \qquad (1)$$

Using two absolute rotations $(R, R') \in SO(3) \times SO(3)$ to represent essential matrix is clearly an over-parametrization, because the essential manifold has only five dofs. The excess one dof can further be removed, as we will show next.

---

[1] Note that, the second camera's center can be set on either X-, Y-, or Z-axis; the resultant parametrization using X- or Y-axis can be similarly derived. We opt for Z-axis for the convenience of closed-form bounding function evaluation ($cf.$ Sec. 4.3).
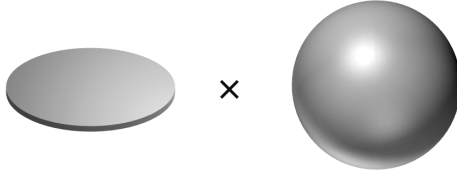
Fig. 1. The essential manifold is parameterized as the product space of a solid 2D disk $\mathbb{D}^2_\pi$ and a solid 3D ball $\mathbb{B}^3_\pi$, corresponding to rotations of the first and second camera respectively. (Note that the disk is thickened to aid in visualization)

Observe that, under our special camera setup, any rotation about Z-axis (*i.e.* the axis joining the two camera centers) applied to both cameras will leave the essential matrix invariant. In other words, they form an *equivalence class* which is a member of the 2D rotation group SO(2). In order to "factor out" these Z-axis rotations, we apply *group quotient operator* to one of the two SO(3) groups as SO(3)/SO(2). In this way we can represent the essential space as SO(3) × (SO(3)/SO(2)), i.e. the product space of SO(3) – rotation space for one camera, and SO(3)/SO(2) for the other camera. Note that there are still equivalence classes remaining, and each of them corresponds to four relative pose configurations [12,30]. It is necessary to leave these equivalence classes there, as only one (unknown) configuration out of the four depicts the true relative pose.

We adopt the angle-axis representation for 3D rotations, with which any rotation is representable as a point in a solid radius-π ball in 3-space, *i.e.* $\mathbb{B}^3_\pi$. Thus SO(3) can be parameterized as $\mathbb{B}^3_\pi$. The remaining problem is how to parameterize SO(3)/SO(2). It is known in topology [17] that SO(3)/SO(2) is homeomorphic to $\mathbb{S}^2$. Instead of this, we directly parameterize SO(3)/SO(2) using angle-axis representation of camera rotation, as detailed in the following.

With angle-axis representation, it is easy to verify that, in our setup the X-Y plane of $\mathbb{B}^3_\pi$ effectively encodes all "Z-axis-free" rotations we need. This is because the X-Y plane of $\mathbb{B}^3_\pi$ contains all rotations whose Z-axis components are zero, while X-axis and Z-axis components are arbitrary. Concretely, let $\mathbf{v}$ be the angle-axis vector of R, *i.e.* $R = \exp([\,\mathbf{v}\,]_\times)$, we avoid the freedom of Z-axis rotation by setting v^3, the 3rd element of vector $\mathbf{v}$, to be 0. Thus our search space for the first rotation $R = \exp([v^1, v^2, 0]_\times)$ is reduced to the 2D disk $\mathbb{D}^2_\pi$ on the equator plane of the π-ball. Now, we have "squeezed" a 3D radius-π ball to a flat 2D radius-π disk in the X-Y plane.

Without loss of generality, we assume the first camera's rotation R is of 2-dof and "Z-axis free"; we denote this as $\mathbf{v} \in \mathbb{D}^2_\pi$. Let $R' = \exp([\,\mathbf{v}'\,]_\times)$, then the essential manifold is parameterized by 5D vectors $(\mathbf{v}, \mathbf{v}') \in \mathbb{D}^2_\pi \times \mathbb{B}^3_\pi$. To recover a 3×3 essential matrix E from $(\mathbf{v}, \mathbf{v}')$, one simply needs to recover rotations matrices (R, R') from $(\mathbf{v}, \mathbf{v}')$, then compute E with Eq. (1).

Comparison to previous work. Some previous works such as [14,29] base their parametrization on Singular Value Decomposition (SVD) of essential matrix. Although these representations also originate from SO(3) × SO(3), they do not provide the geometric interpretation of their parameters, and are not suitable for our BnB search. Very recently, an independent work [30] chooses the same

coordinate system as ours and uses the essential matrix formulation in Eq. (1). One difference between [30] and our work is that, [30] computes geodesic distance between two equivalence classes of two 6D $\mathrm{SO}(3) \times \mathrm{SO}(3)$ elements, while we propose an explicit parametrization of the 5D manifold $\mathrm{SO}(3) \times \left(\mathrm{SO}(3)/\mathrm{SO}(2)\right)$.

3   Optimization Criteria

With the parametrization described above, we are ready to formally define the optimality, and formulate the problem we will solve.

Let $(\mathbf{x}, \mathbf{x}')$ be a putative feature correspondence pair represented as unit 3D vectors, both corresponding to an unknown 3D scene point $\mathbf{X} \in \mathbb{R}^3$. Note $(\mathbf{x}, \mathbf{x}')$ may be subject to outliers and measurement noise. We represent the two cameras by their absolute orientations R and R', which jointly encode the essential matrix $\mathrm{E} = \mathrm{E}(\mathrm{R}, \mathrm{R}')$. The epipolar equation $\mathbf{x}'^\mathrm{T}\mathrm{E}\mathbf{x} = 0$ gives an algebraic error metric for measuring the optimality of an essential matrix. In this work, we will use the geometrically meaningful angular reprojection error, which is defined as

$$
\begin{aligned}
\angle(\mathrm{R}^\mathrm{T}\mathbf{x}, \mathrm{R}'^\mathrm{T}\mathbf{x}') &\doteq \min_{\mathbf{X}} \max\left(\angle(\mathrm{R}^\mathrm{T}\mathbf{x}, \mathbf{X}), \angle(\mathrm{R}'^\mathrm{T}\mathbf{x}', \mathbf{X} - \mathbf{C}')\right) \\
&= \min_{\mathbf{X}} \max\left(\angle(\mathbf{x}, \mathrm{R}\mathbf{X}), \angle(\mathbf{x}', \mathrm{R}'(\mathbf{X} - \mathbf{C}'))\right)
\end{aligned}
\tag{2}
$$

where $\angle(\cdot, \cdot)$ denotes the angle between two vectors, and $\mathbf{C}' \equiv [0, 0, 1]^\mathrm{T}$. We use the symbol $\angle(\cdot, \cdot)$ to denote the angular reprojection error, which is the maximum of the two angular residuals.

With this angular error definition, there are two options to define the optimality of essential matrix $\mathrm{E}(\mathrm{R}, \mathrm{R}')$, corresponding to the following two problems.

Problem 1 (Inlier-set cardinality maximization). Given feature correspondences $(\mathbf{x}_i, \mathbf{x}'_i)$ and a prescribed angular error tolerance ϵ, the optimal essential matrix $\mathrm{E}(\mathrm{R}, \mathrm{R}')$ maximizes the cardinality of the inlier set (or consensus set) as

$$
\max_{\mathrm{R}, \mathrm{R}'} |\mathcal{I}|, \quad s.t. \ \forall i \in \mathcal{I}, \ \angle(\mathrm{R}^\mathrm{T}\mathbf{x}_i, \mathrm{R}'^\mathrm{T}\mathbf{x}'_i) \leq \epsilon
\tag{3}
$$

where $\mathcal{I}$ denotes the inlier set and $|\cdot|$ represents cardinality. A pair of correspondences $(\mathbf{x}_i, \mathbf{x}'_i)$ is considered to be an inlier *w.r.t.* ϵ if $\angle(\mathrm{R}^\mathrm{T}\mathbf{x}_i, \mathrm{R}'^\mathrm{T}\mathbf{x}'_i) \leq \epsilon$.

Problem 2 (Angular reprojection error minimization). Given feature correspondences $(\mathbf{x}_i, \mathbf{x}'_i)$, the optimal essential matrix $\mathrm{E}(\mathrm{R}, \mathrm{R}')$ is found by

$$
\min_{\mathrm{R}, \mathrm{R}'} \|\mathbf{e}\|, \quad s.t. \ e_i = \angle(\mathrm{R}^\mathrm{T}\mathbf{x}_i, \mathrm{R}'^\mathrm{T}\mathbf{x}'_i)
\tag{4}
$$

where $\|\cdot\|$ is a certain norm.

Solving Problem 2 gives rise to an exact essential matrix minimizing angular error; however the result is sensitive to outliers. The goal of this paper is to optimally solve Problem 1 with an exact inlier-set cardinality, thus it is intrinsically

robust. Note that the solution to Problem 1 may not be unique. To solve essential matrix both robustly and exactly, one can solve Problem 2 with existing methods (*e.g.* [11]) after obtaining the true inliers with the proposed method.

Although global optimization for Problem 2 is studied in [11], solving the cardinality maximization problem globally optimally is still extremely difficult due to its obvious combinatorial and discrete nature. In this paper, we approach the problem as a continuous optimization, and solve it by BnB search over the continuous parameter domain – the 5D product space $\mathbb{D}_\pi^2 \times \mathbb{B}_\pi^3$.

4   Branch and Bound over $\mathbb{D}_\pi^2 \times \mathbb{B}_\pi^3$

Recall that the goal is to globally maximize the inlier-cardinality as shown in Eq. (3). We treat this problem as continuous optimization, and solve it via 5D space BnB. A high level description of our method is given below. For the ease of manipulation, we use a 5D cube $\mathbb{C}_\pi^5$ with half side-length π to enclose the space of $\mathbb{D}_\pi^2 \times \mathbb{B}_\pi^3$. The initial cube $\mathbb{C}_\pi^5$ can be divided into smaller cubes. For each such cube, we compute the lower-bound (*LB*) as well as the upper-bound (*UB*) of the inlier-set cardinality for all rotations within it. *LB* and *UB* will be compared with the best value found so far, then this cube will be discarded or sub-divided. In the following we will denote a cube by $C_\sigma(\bar{\mathrm{R}}, \bar{\mathrm{R}}')$, where σ is its half side-length, and $\bar{\mathrm{R}}, \bar{\mathrm{R}}'$ are the center rotations of the corresponding 2D square and 3D cube respectively.

As is true for any BnB algorithm, the key to success is to find effective and efficient bounds. Below we will explain how we achieve this.

4.1   Lower-bound computation

Finding a lower-bound for the cardinality maximization problem is relatively easy. It can be done simply by evaluating the cardinality function at a single point within the cubical domain. Obviously, the cardinality obtained in this way is necessarily a lower-bound, as it must not be greater than the true maximal cardinality with rotation in that cube.

The following algorithm computes a lower-bound for a cube $C_\sigma(\bar{\mathrm{R}}, \bar{\mathrm{R}}')$ *w.r.t.* a prescribed angular error tolerance ϵ.
1. Check all candidate correspondences $(\mathbf{x}_i, \mathbf{x}_i')$, with center rotations $\bar{\mathrm{R}}, \bar{\mathrm{R}}'$.
2. Count how many *feasibility inequalities* $\angle(\mathrm{R}^\mathrm{T}\mathbf{x}_i, \mathbf{X}_i) \le \epsilon$ and $\angle(\mathrm{R}'^\mathrm{T}\mathbf{x}_i', \mathbf{X}_i - \mathbf{C}') \le \epsilon$ can be satisfied with some $\mathbf{X}_i$.
3. Report the above count as a lower-bound for this cube.

Step 2 of the algorithm is done by solving a series of *feasibility test problems*. How to perform such tests will be explained in Sec. 4.3.

4.2   Upper-bound computation via relaxation

In solving maximization (as opposed to minimization) with BnB, it is in general more difficult to find a proper upper-bound (than to find a lower-bound).

The following algorithm gives our solution to finding suitable upper-bound of the cardinality function for a given cube $C_\sigma(\bar{R}, \bar{R}')$ and tolerance ϵ.

1. Check all correspondences $(\mathbf{x}_i, \mathbf{x}'_i)$ with center rotations $\bar{R}, \bar{R}'$.
2. Count how many *relaxed feasibility inequalities* $\angle(\bar{R}^T \mathbf{x}_i, \mathbf{X}_i) \leq \epsilon + \sqrt{2}\sigma$ and $\angle(\bar{R}'^T \mathbf{x}'_i, \mathbf{X}_i - \mathbf{C}') \leq \epsilon + \sqrt{3}\sigma$ can be satisfied with some $\mathbf{X}_i$.
3. Report the above count as an upper-bound for this cube.

Note in Step 2 of this algorithm, we solve a *relaxed* feasibility test problem, as the thresholds in the right side of the inequalities have been enlarged (relaxed), leading to more correspondences to be claimed as inliers, hence increasing the inlier cardinality.

To show that the upper-bound is valid (*i.e.* no solution in the cube yields larger inlier-set cardinality), a lemma and its proof are given below.

Lemma 1. *For a 5D cubic domain* $C_\sigma(\bar{R}, \bar{R}')$, *solving the above relaxed feasibility problem gives a valid upper-bound of the inlier-set cardinality.*

Proof. Our proof follows from two lemmas of paper [11], which show that, for any vector $\mathbf{x} \in \mathbb{R}^3$, given two arbitrary rotations $R, \bar{R}$ (with $\mathbf{v}$ and $\bar{\mathbf{v}}$ as their angle-axis representations), one must have $\angle(R\mathbf{x}, \bar{R}\mathbf{x}) \leq \angle(R, \bar{R}) \leq \|\mathbf{v} - \bar{\mathbf{v}}\|$.

Let's first fix R', and consider a 2D square domain of R centered at $\bar{R}$ with half side-length σ. Suppose R^* is the optimal rotation, among all rotations within this domain, such that the corresponding inlier-set $\mathcal{I}$ is maximized. Therefore R^* must be feasible for inlier points, *i.e.* $\forall i \in \mathcal{I}$ one has $\angle(R^{*T}\mathbf{x}_i, R'^T\mathbf{x}'_i) \leq \epsilon \Rightarrow \angle(R^{*T}\mathbf{x}_i, \mathbf{X}_i) \leq \epsilon$ with some $\mathbf{X}_i$. Then for the center rotation $\bar{R}$ we have

$$\begin{aligned}
\angle(\bar{R}^T\mathbf{x}_i, \mathbf{X}_i) &\leq \angle(R^{*T}\mathbf{x}_i, \mathbf{X}_i) + \angle(\bar{R}^T\mathbf{x}_i, R^{*T}\mathbf{x}_i) \\
&\leq \epsilon + \angle(\bar{R}, R^*) \\
&\leq \epsilon + \|\bar{\mathbf{v}} - \mathbf{v}^*\| \\
&\leq \epsilon + \sqrt{2}\sigma.
\end{aligned} \tag{5}$$

This result implies that, if we relax the right side of the feasibility inequality from ϵ to $\epsilon + \sqrt{2}\sigma$ and evaluate inlier cardinality with respect to the center rotation, then the obtained cardinality will be no less than the optimal cardinality obtained within this cube, *i.e.* the one corresponding to R^*.

For the other rotation R' (which is a 3-dof rotation) and vector $\mathbf{v}'$, a similar result can be obtained, except that in this case one has $\sqrt{3}\sigma$ for a 3D cubic domain instead of $\sqrt{2}\sigma$. Combining both rotations we have: for each point i in the optimal inlier-set with rotations in $C_\sigma(\bar{R}, \bar{R}')$, both $\angle(\bar{R}^T\mathbf{x}_i, \mathbf{X}_i) \leq \epsilon + \sqrt{2}\sigma$ and $\angle(\bar{R}'^T\mathbf{x}'_i, \mathbf{X}_i - \mathbf{C}') \leq \epsilon + \sqrt{3}\sigma$ must be satisfied with some $\mathbf{X}_i$. This completes the proof and the upper-bound is valid.

4.3   Efficient bounding with closed-form feasibility test

Solving upper-bound and lower-bound necessitates the feasibility test task. This task is: given a pair of camera rotations R, R' (along with $\mathbf{C} \equiv \mathbf{0}, \mathbf{C}' \equiv [0, 0, 1]^T$), test whether or not a correspondences pair $(\mathbf{x}, \mathbf{x}')$ is an inlier *w.r.t.* the given angular reprojection error threshold ϵ. It can be formally formulated as

Problem 3 (Feasibility test for determining inliers).

$$
\begin{aligned}
&\text{Given} && \mathbf{x}, \mathbf{x}', R, R', \mathbf{C}, \mathbf{C}', \varepsilon, \varepsilon' \\
&\text{does there exist } \mathbf{X} && \\
&\text{such that} && \angle(R^{\mathrm{T}}\mathbf{x}, \mathbf{X} - \mathbf{C}) \leq \varepsilon \\
&\text{and} && \angle(R'^{\mathrm{T}}\mathbf{x}', \mathbf{X} - \mathbf{C}') \leq \varepsilon'
\end{aligned}
$$

where $\varepsilon = \varepsilon' = \epsilon$ for the feasibility test in lower-bound computation, and $\varepsilon = \epsilon + \sqrt{2}\sigma$, $\varepsilon' = \epsilon + \sqrt{3}\sigma$ for the relaxed one in upper-bound computation.

One way to do such a test is by *two-view triangulation*. It has been shown in [16,15] that this problem can be solved by Second Order Cone Programming (SOCP). We have tested this method experimentally using a commercial SOCP solver (MOSEK). It worked successfully on very small numbers of feature points but with high computational demand, preventing us from doing larger experiment. We were therefore motivated to seek a faster solution.

In this paper, built upon previous work [5], our bounds are derived with efficient feasibility test in closed-form. The intuition is: to verify whether or not $(\mathbf{x}, \mathbf{x}')$ is compatible with a tentatively given essential matrix, one does not have to recover the corresponding 3D point $\mathbf{X}$. Instead, it is sufficient to check whether or not the epipolar relationship of the two points is satisfied. A similar idea was proposed in [11], where a Linear Programming solver is used for feasibility tests.

Our method avoids using convex programming. It is a direct application of the following theorem which is a simple extension of that in [5]. Recall that, the first camera is centered the origin and the second one is on Z-axis. If we represent the unit vectors $R^{\mathrm{T}}\mathbf{x}$ and $R'^{\mathrm{T}}\mathbf{x}'$ in spherical coordinates, they become

$$
R^{\mathrm{T}}\mathbf{x} = \begin{bmatrix} \sin\theta\cos\varphi \\ \sin\theta\sin\varphi \\ \cos\theta \end{bmatrix}, \quad R'^{\mathrm{T}}\mathbf{x}' = \begin{bmatrix} \sin\theta'\cos\varphi' \\ \sin\theta'\sin\varphi' \\ \cos\theta' \end{bmatrix}. \tag{6}
$$

Theorem 1. *Given a pair of correspondences $\mathbf{x}, \mathbf{x}'$, rotation matrices R, R' and camera centers $\mathbf{C} \equiv \mathbf{0}, \mathbf{C}' \equiv [0,0,1]^{\mathrm{T}}$, representing $R^{\mathrm{T}}\mathbf{x}$ and $R'^{\mathrm{T}}\mathbf{x}'$ in spherical coordinates as (θ, φ) and (θ', φ'), we have: Problem 3 is feasible if and only if*
$$
\begin{cases} \theta \leq \theta' + \varepsilon + \varepsilon' \\ |\varphi - \varphi'| \leq \omega \end{cases}, \text{ where } \omega \text{ is given below}
$$

$$
\omega = \begin{cases} \arcsin(\frac{\sin\varepsilon}{\sin\theta}) + \arcsin(\frac{\sin\varepsilon'}{\sin\theta'}), & if\ \theta < \theta' \\ \arccos(\frac{\cos(\varepsilon+\varepsilon') - \cos\theta\cos\theta'}{\sin\theta\sin\theta'}), & if\ \theta \in [\theta', \theta' + \varepsilon + \varepsilon'] \\ \pi, & if\ any\ of\ the\ above\ is\ undefined \end{cases} \tag{7}
$$

Proof of this theorem can be found in [5]. The geometric intuition behind Theorem 1 is easy to discern. Consider the limit case when $\varepsilon \to 0$ and $\varepsilon' \to 0$ (thus $\omega \to 0$), then $|\varphi - \varphi'| \leq \omega \Rightarrow \varphi = \varphi'$ says that the two viewing rays of the two points lie in the same *half-plane* containing the baseline, and $\theta < \theta' + \varepsilon + \varepsilon' \Rightarrow \theta < \theta'$ entails that the two viewing rays intersect in this half-plane.

Based on this theorem, both lower-bound and upper-bound for a cube can be made in closed-form. The evaluation is efficient with elementary computation (and counting), using basic trigonometric functions.

Algorithm 1: BnB search in $\mathbb{D}_\pi^2 \times \mathbb{B}_\pi^3$ for optimal essential matrix maximizing the inlier set

---

    Input: Images point pairs $(\mathbf{x}_i, \mathbf{x}_i'), i = 1, \ldots, M$; angular error threshold ϵ.
    Output: Optimal essential matrix E^* and corresponding inlier set $\mathcal{I}^*$ of size N^*.

1 Divide $[-\pi, \pi]^5$ into small sub-cubes and push them into priority queue Q.
2 Set $N^* = 4$.   *%we need to find at least $N^* = 5$ points*
3 begin
4   Read out a cube with the highest upper-bound UB from Q.
5   Quit the loop if $UB = N^*$.
6   Divide it into $2^5 = 32$ sub-cubes with equal side length.
7   **foreach** sub-cube $C_\sigma(\bar{R}, \bar{R}')$ **do**
8   Set its lower-bound LB and upper-bound UB to be 0.
9   **foreach** correspondence pair $(\mathbf{x}_i, \mathbf{x}_i')$ **do**
10 LB++, if Problem 3 is feasible with $\bar{R}, \bar{R}', \epsilon, \epsilon$.
11 UB++, if Problem 3 is feasible with $\bar{R}, \bar{R}', \epsilon + \sqrt{2}\sigma, \epsilon + \sqrt{3}\sigma$.
12   **end**
13   **if** $LB > N^*$ **then**
14   Update $N^* = LB$, $E^* = E(\bar{R}, \bar{R}')$ and also $\mathcal{I}^*$.
15   **end**
16   Discard this cube if $UB \leqslant N^*$; otherwise put it into Q.
17   **end**
18 end

Degeneracy. Note that when a feature point (θ, φ) either falls on Z-axis or is sufficiently close to it ($\theta < \epsilon$ or $\theta < \epsilon'$), the above functions for ω are not defined. In such cases the feasibility test always returns true.

4.4 The main algorithm

Armed with the above developments of domain parametrization, lower and upper bounds, and closed-form feasibility test, we are now ready to present our main algorithm. Although it appears to be a bit technically heavy, the central idea and the implementation are rather simple: for each parameter domain, *i.e.* a 5D cube, count the number of feature correspondences that pass the feasibility test (or, relaxed feasibility test) as the lower-bound (or, upper-bound) of the cardinality, and try to update the solution and discard this cube accordingly. Algorithm 1 summarizes the algorithm in pseudo-code form.

Initial cubes. Before the BnB loop we divide the initial cube $[-\pi, \pi]^5$ into smaller cubes as it is less likely that a large cube can be discarded. In our implementation we use $6^5 = 7776$ initial cubes with equal side length.

Search strategy. The BnB algorithm uses the *best-first-search* strategy. Concretely, it maintains a priority queue of the active cubes, whose priorities are set to be their upper-bounds. In this way, the BnB algorithm always explores the most promising cube first.
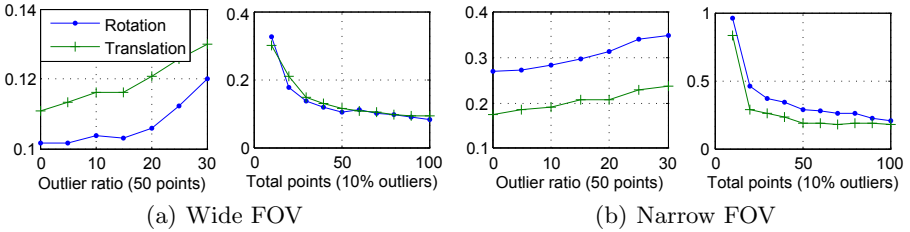
Fig. 2. Average rotation and translation error (both in degrees) for 50 runs of our method in synthetic wide-FOV (*left*) and narrow-FOV (*right*) tests *w.r.t.* different outlier ratios and total points.
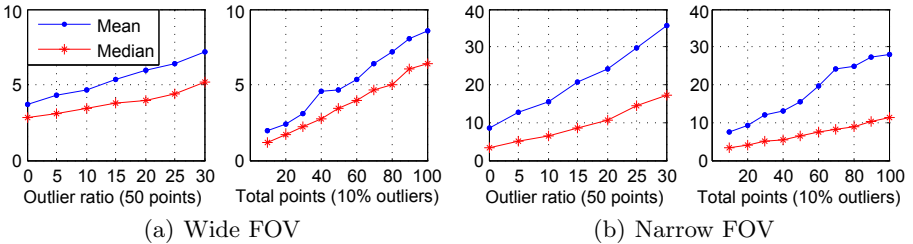

Fig. 3. Running time (in seconds) for 50 runs of our method in synthetic wide-FOV (*left*) and narrow-FOV (*right*) tests *w.r.t.* different outlier ratios and total points.

Proof of convergence. The convergence of the algorithm is easy to see, as when the side-lengths of all cubes asymptotically diminish to zero, the gap between the upper-bound and lower-bound will be zero too.

5   Experiment Results

In this section, we report the experimental results on synthetic scenes and real imageries. Our method is implemented in C++, and tested on a standard PC with Intel i7 3.4GHz 4-core CPU and 8GB memory.

5.1   Synthetic scene test: normal cases

The main goal of experiments on synthetic data is to verify the correctness of the proposed method, including the essential manifold parametrization and the BnB algorithm. In these experiments we set the angular error threshold to be 0.002 radians (about 0.115 degrees). Inlier number is the main index for essential matrix evaluation as our goal is to optimally maximize it. Nevertheless, we will also report the estimation error of essential matrix. For better comprehension, we use classic parametrization $E = [\hat{\mathbf{t}}]_\times \hat{R}$, and evaluate error of $\hat{R}$ and $\hat{\mathbf{t}}$. Rotation error is the angle between $\hat{R}$ and ground truth rotation. As $\hat{\mathbf{t}}$ is obtained

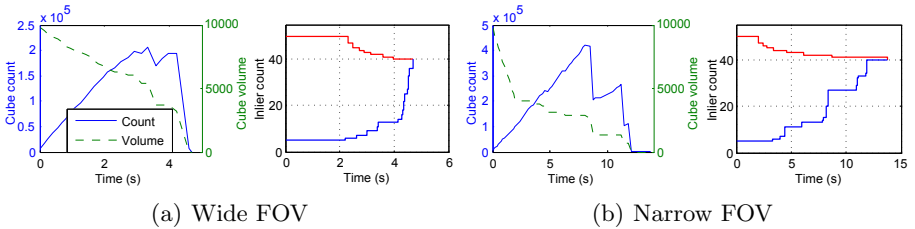(a) Wide FOV   (b) Narrow FOV

Fig. 4. Typical cube and bound evolutions of BnB in synthetic wide-FOV (*left*) and narrow-FOV (*right*) tests using 50 points with 20% (*i.e.* 10) outliers.

up to a scale, we define translation error as the angle between $\hat{\mathbf{t}}$ and ground truth translation. Note that, as discussed in Sec. 3, these results can be further improved by minimizing the reprojection error of obtained inliers (which is not used here).

Wide Field-Of-View (Omnidirectional Camera). In this test synthetic data with random points and two omnidirectional cameras which have 360° field of view were used. We synthesized 50 configurations of different points and camera poses. The points were generated in a cube centered at origin with side length 4, and camera centers were generated from Gaussian distribution centered at origin with $\sigma = 0.5$. Gaussian noise with $\sigma = 0.001$ was added to all the projected image points. To generate outliers, we randomly perturbed the image points in the first camera by over 10 degrees. We tested our method first on different numbers of outliers with fixed total points (50), and then on different numbers of points with fixed outlier ratio (10%). As expected, *our method succeeded in all the tests in terms of finding out all the true inliers.* Average rotation and translation errors of the 50 configurations are shown in Fig. 2. Clearly, the error increases with outlier ratio, and decreases with total point number. Average running time is shown in Fig. 3. In general, it took the method longer time to converge when higher levels of outliers were present. To visualize the behavior of BnB, we present typical evolution curves of active cubes and global bounds as a function of time in Fig. 4.

Narrow Field-Of-View. We then tested our method with narrow field of view. We synthesized the situation where the points are confined in approximately 60° FOV of two regular pinhole cameras. The points were generated in a cube centered at origin with side length 4, and cameras were randomly placed at a distance of about 4 facing the origin. Other settings were the same with that in wide-FOV tests. Again, our method successfully found out all the true inliers. The estimation error, running time, and typical BnB evolution are also shown in Fig. 2, Fig. 3 and Fig. 4 respectively. It is clear that solving the problem with narrow-FOV is generally more difficult than that with wide-FOV, as evidenced by the larger rotation and translation errors as well as the longer running time of our method.

5.2   Synthetic scene test: special cases

We tested some special cases on wide FOV configuration, aiming to test the performance of the proposed method under special or extreme situations.

Large outlier ratio. To test the performance under large outlier ratio, we generated 50 points with 25 (50%), 30 (60%), 35 (70%) outliers respectively in the wide FOV configuration. Our method successfully found the true inliers in 11s, 26s and 81s respectively.

Pure translational motion. In this experiment, two cameras with pure (and random) translation as the ground-truth transformation and 50 points were synthesized. We ran our method on these points, and the angle between the two estimated rotations R and R′ is about 0.11 degrees, which indicates that our method successfully identified the equal rotation case.

All scene points on a plane. We synthesized a planar case where all 50 points lie on a plane. This is a well-known degenerate case for fundamental matrix estimation, however it should not affect essential matrix estimation, as explained in [24]. Our experiment on this case obtained positive result and we successfully recovered the correct essential matrices with and without outliers. The rotation and translation errors are all below 0.15 degrees.

5.3   Real image test

Images from both narrow-FOV and wide-FOV cameras were then used to evaluate the real-life performance of our method. We also tested RANSAC and LO-RANSAC [2] (with Option 4 of local optimization described in [2]) methods. In both RANSAC implementations, the 8-point method[2] was used and angular error threshold is adopted to distinguish outliers; the outlier ratio and probability parameter η were set to be 30% and 0.99 respectively. Note that, the goal of this paper is not to replace the popular RANSAC and its variants in essential matrix estimation, but to provide a complementary (yet important) optimal method.

Narrow Field-Of-View. We tested our method on two image pairs from the Corridor and Valbonne data sets[3]. 94 and 106 SIFT matches [22] were generated respectively for the two pairs as shown in Fig. 5. The angular error threshold was set to 0.0015 radians. We parallelized the BnB search with 8 threads, and our method converged in 221s and 453s respectively. Apparently, it takes quite more time than on synthetic data of the same size. However, this is reasonable as will be analyzed as follows. On a 600×600 image from a 60°-FOV camera, a small pixel difference, say 3 pixels, yields about 0.3-degree angle difference. To tell outliers from inliers at this accuracy of both camera orientations, the 5D cube would have to be divided into $(\frac{180}{0.3/\sqrt{2}})^2 \times (\frac{180}{0.3/\sqrt{3}})^3 \approx 8\times10^{14}$ blocks for a complete search method, and this is also a very difficult task for our BnB. The number

---

[2] The 5-point method (with [9]'s solver) was also tested. It performed comparably with or slightly worse than the 8-point method; the latter one is thus presented.

[3] http://www.robots.ox.ac.uk/~vgg/data/data-mview.html

Table 1. Inlier-set maximization performance of different methods. The first column lists the images and correspondence numbers. The second and third columns show the maximal and mean inlier number detected by RANSAC and LO-RANSAC in their 1,000 runs. The last column shows the inlier number from our method and the running time (with 8 threads).

Images (#points)	RANSAC max/mean #inliers	LO-RANSAC max/mean #inliers	Our method #inliers (time)
Corridor (94)	63 / 32.5	65 / 50.0	66 (221s)
Church (106)	82 / 32.8	87 / 35.5	89 (453s)
Building (202)	160 / 146.9	161 / 151.7	163 (52s)
Office (151)	124 / 91.4	126 / 104.2	126 (43s)

of detected inliers is 66 for Corridor image and 89 for Church image, indicating 29.8% and 16% outlier ratios respectively. The detected inliers and outliers are shown in Fig. 5. For some outliers we show their angular errors (optimally solved via bi-section and SOCP [15]) with the obtained essential matrix.

We then repeated both RANSAC and LO-RANSAC 1,000 times with the same angular error threshold; the resulting inlier numbers are shown in the first two columns of Tab. 1. The heuristic and stochastic nature of random sampling scheme can be clearly seen, as the mean performances of the 1,000 runs are not satisfactory. Moreover, both the two methods failed to detect the same inlier number as ours. This can be explained by the fact that algebraic solution of essential matrix is not consistent with the meaningful geometric error metric. In future we plan to compare our method with RANSAC methods in high-noise situation where algebraic solutions can be severely biased.

Wide Field-Of-View (Fisheye Camera). In order to test our method in real-life wide-FOV case, a camera with a fisheye lens was used to capture images of the scene with up to 190° FOV. The camera was calibrated with the method of [27].

Fig. 6 shows two typical pairs referred as Building and Office. The angular threshold was set to be 0.003 radians for these images. Our method converged in 52s and 43s for the two image pairs respectively as shown in Tab. 1, and the results indicate 19.3% and 16.6% outlier ratios. In general, the angular errors of outliers are larger than that in the narrow-FOV case (see Fig. 6), and our method ran faster on wide-FOV images. This result is in consistent with our synthetic experiments and similar discoveries reported in previous works [3,11,6,13].

6   Conclusion and Future Work

A branch-and-bound global optimization method is proposed for essential matrix estimation via inlier-set cardinality maximization under geometric (angular) error. An explicit and geometrically meaningful parametrization of the 5D essential manifold, *i.e.* $\mathbb{D}^2_\pi \times \mathbb{B}^3_\pi$, is used to perform the BnB search. Based on previous works [11] and [5], closed-form bounding functions of inlier-set cardinality are derived, leading to efficient bound evaluation in the 5D space BnB.
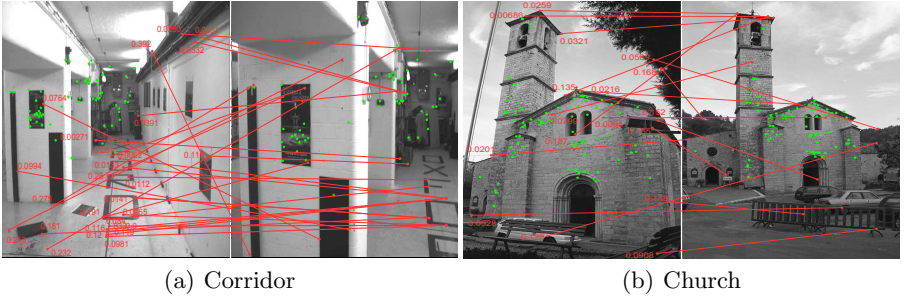
(a) Corridor   (b) Church

Fig. 5. Results on narrow-FOV images. Green and red dots are respectively inlier and outlier correspondences found by our method. For outliers we labeled their angular reprojection errors in radius. (Best viewed on screen and with zoom-in)
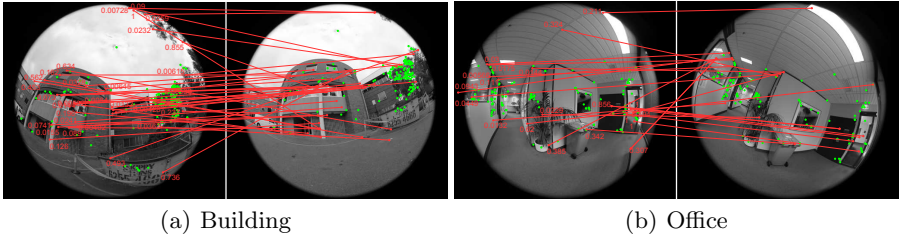

(a) Building   (b) Office

Fig. 6. Results on wide-FOV images taken with a fisheye camera. See caption of Fig. 5.

Currently the proposed method is slow especially for cameras with small field of view. Nevertheless, due to its optimality the method can be used as a benchmark for method evaluation, or be applied in situations where robustness or accuracy is highly desired while speed is not crucial.

To make the method faster and more practical, there are some strategies we would like to investigate in future. For example, a possible one is to get an initial essential matrix estimate using RANSAC, then search the parameter space with the proposed BnB in a small region around this estimate. Taking advantage of prior knowledge on motion to confine the parameter space is a metric of continues optimization in contrast to discrete combinatorial optimization. Since our BnB algorithm can be easily parallelized, another idea would be porting it onto modern GPU where a significant speedup can be expected.

Acknowledgements

References

1. Bazin, J.C., Seo, Y., Pollefeys, M.: Globally optimal consensus set maximization through rotation search. In: ACCV (2012)
2. Chum, O., Matas, J., Kittler, J.: Locally optimized RANSAC. Pattern Recognition (2003)
3. Daniilidisl, K., Spetsakisz, M.E.: Understanding noise sensitivity in structure from motion. In: Visual navigation: from biological systems to unmanned ground vehicles. Psychology Press (1997)
4. Dellaert, F., Seitz, S.M., Thorpe, C.E., Thrun, S.: Structure from motion without correspondence. In: CVPR (2000)
5. Enqvist, O., Jiang, F., Kahl, F.: A brute-force algorithm for reconstructing a scene from two projections. In: CVPR (2011)
6. Enqvist, O., Kahl, F.: Two view geometry estimation with outliers. In: BMVC (2009)
7. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm. ACM (1981)
8. Goshen, L., Shimshoni, I.: Balanced exploration and exploitation model search for efficient epipolar geometry estimation. T-PAMI (2008)
9. Hartley, R., Li, H.: An efficient hidden variable approach to minimal-case camera motion estimation. T-PAMI (2012)
10. Hartley, R.I.: In defense of the eight-point algorithm. T-PAMI (1997)
11. Hartley, R.I., Kahl, F.: Global optimization through searching rotation space and optimal estimation of the essential matrix. In: ICCV (2007)
12. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision (2nd Edition). Cambridge University Press (2004)
13. Heller, J., Havlena, M., Pajdla, T.: A branch-and-bound algorithm for globally optimal hand-eye calibration. In: CVPR (2012)
14. Helmke, U., Hüper, K., Lee, P.Y., Moore, J.: Essential matrix estimation using gauss-newton iterations on a manifold. IJCV (2007)
15. Kahl, F., Hartley, R.: Multiple-view geometry under the l_∞-norm. T-PAMI (2008)
16. Ke, Q., Kanade, T.: Quasiconvex optimization for robust geometric reconstruction. ICCV (2005)
17. Lee, J.: Introduction to topological manifolds. Springer (2010)
18. Li, H.: A practical algorithm for l_∞ triangulation with outliers. In: CVPR (2007)
19. Li, H.: Consensus set maximization with guaranteed global optimality for robust geometry estimation. In: ICCV (2009)
20. Li, H., Hartley, R.: Five-point motion estimation made easy. In: ICPR (2006)
21. Longuet-Higgins, H.C.: A computer algorithm for reconstructing a scene from two projections. Nature (1981)
22. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV (2004)
23. Makadia, A., Geyer, C., Daniilidis, K.: Correspondence-free structure from motion. IJCV (2007)
24. Nistér, D.: An efficient solution to the five-point relative pose problem. T-PAMI (2004)
25. Olsson, C., Eriksson, A., Hartley, R.: Outlier removal using duality. In: CVPR (2010)
26. Raguram, R., Chum, O., Pollefeys, M., Matas, J., Frahm, J.: USAC: A universal framework for random sample consensus. T-PAMI (2013)

27. Scaramuzza, D., Martinelli, A., Siegwart, R.: A toolbox for easily calibrating omnidirectional cameras. In: IROS (2006)
28. Sim, K., Hartley, R.: Removing outliers using the l_∞ norm. In: CVPR (2006)
29. Subbarao, R., Genc, Y., Meer, P.: Robust unambiguous parametrization of the essential manifold. In: CVPR (2008)
30. Tron, R., Daniilidis, K.: On the quotient representation for the essential manifold. In: CVPR (2014)
31. Yang, J., Li, H., Jia, Y.: Go-ICP: Solving 3d registration efficiently and globally optimally. In: ICCV (2013)
32. Zheng, Y., Sugimoto, S., Okutomi, M.: A branch and contract algorithm for globally optimal fundamental matrix estimation. In: CVPR (2011)