Globally-Optimal Inlier Set Maximisation for Simultaneous Camera Pose and Feature Correspondence

Dylan Campbell[1,2], Lars Petersson[1,2], Laurent Kneip[1] and Hongdong Li[1]
[1]Australian National University* [2]Data61 – CSIRO
{dylan.campbell,lars.petersson,laurent.kneip,hongdong.li}@anu.edu.au

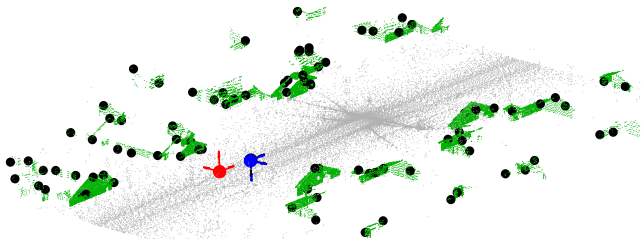
Abstract

Estimating the 6-DoF pose of a camera from a single image relative to a pre-computed 3D point-set is an important task for many computer vision applications. Perspective-n-Point (PnP) solvers are routinely used for camera pose estimation, provided that a good quality set of 2D–3D feature correspondences are known beforehand. However, finding optimal correspondences between 2D key-points and a 3D point-set is non-trivial, especially when only geometric (position) information is known. Existing approaches to the simultaneous pose and correspondence problem use local optimisation*, and are therefore unlikely to find the optimal solution without a good pose initialisation, or introduce restrictive assumptions. Since a large proportion of outliers are common for this problem, we instead propose a globally-optimal inlier set cardinality maximisation approach which jointly estimates optimal camera pose and optimal correspondences. Our approach employs branch-and-bound to search the 6D space of camera poses, guaranteeing global optimality without requiring a pose prior. The geometry of $SE(3)$ is used to find novel upper and lower bounds for the number of inliers and local optimisation is integrated to accelerate convergence. The evaluation empirically supports the optimality proof and shows that the method performs much more robustly than existing approaches, including on a large-scale outdoor data-set.*

1. Introduction

Estimating the pose of a calibrated camera given a set of 2D points in the camera frame and a set of 3D points in the world frame, as shown in Figure 1, is a fundamental part of the general 2D–3D registration problem of aligning an image with a 3D scene or model. When correspondences are known, this becomes the Perspective-n-Point (PnP) problem for which many solutions exist [16, 26, 23, 18, 22]. Applications include camera localisation and tracking [13, 38, 24], augmented reality [34], motion segmentation [39] and object recognition [19, 36, 2].

*This research is supported by an Australian Government Research Training Program (RTP) Scholarship.



(a) 3D point-set (grey and green), 3D features (black dots) and ground-truth (black), RANSAC (red) and our (blue) camera poses. The ground-truth and our camera poses coincide, whereas the RANSAC pose has a translation offset and a 180° rotation offset. Best viewed in colour.

(b) Panoramic photograph and extracted 2D features (top), building points projected onto the image using the RANSAC camera pose (middle) and building points projected using our camera pose (bottom).

Figure 1. Estimating the pose of a calibrated camera from a single image within a large-scale, unorganised 3D point-set captured by vehicle-mounted laser scanner. Our method solves the absolute pose problem while simultaneously finding feature correspondences, using a globally-optimal branch-and-bound approach with tight novel bounds on the cardinality of the inlier set.

While hypothesise-and-test frameworks like RANSAC [13] can mitigate the sensitivity of PnP solvers to outliers in the correspondence set, few approaches are able to handle the case where 2D–3D correspondences are not known in advance. Unknown correspondences arise in many circumstances, including the general case of aligning an image with a textureless 3D point-set or CAD model. While feature extraction techniques provide a relatively robust and reproducible way to detect interest points such as edges or corners within each modality, finding correspondences across the two modalities is much more challenging. Even when the point-set has sufficient visual information associated with it, such as colour or SIFT features [32], repetitive features, occlusions and perspective distortion make the correspondence problem non-trivial. Moreover, appear-

ance and thus visual features may change significantly between viewpoints, lighting conditions, weather and seasons, whereas scene geometry is often less affected. When relocalising a camera in a previously mapped environment or bootstrapping a tracking algorithm, we contend that geometry is often more reliable. Therefore, there is a need for methods that solve for both pose and correspondences.

Efficient local optimisation algorithms for solving this joint problem have been proposed [9, 35]. However, they require a pose prior, search only for local optima and do not provide an optimality guarantee, yielding erroneous pose estimates without a reliable means of detecting failure. Hypothesise-and-test approaches such as RANSAC [13], when applied to the correspondence-free problem [15], are global methods that are not reliant on pose priors but quickly become computationally intractable as the number of points and outliers increase and do not provide an optimality guarantee. More recently, a global and ϵ-suboptimal method has been proposed [5], which uses a branch-and-bound approach to find a camera pose whose trimmed geometric error is within ϵ of the global minimum.

This work is the first to propose a global and optimal inlier set cardinality maximisation solution to the simultaneous pose and correspondence problem. The approach employs the branch-and-bound framework to guarantee global optimality without requiring a pose prior, ensuring that it is not susceptible to local optima. We use a parametrisation of $SE(3)$ space that facilitates branching and derive novel bounds on the objective function. In addition, we also apply local optimisation whenever the algorithm finds a better transformation, to accelerate convergence without voiding the optimality guarantee. Cardinality maximisation allows an exact optimiser to be found, unlike the ϵ-suboptimality inherent to the continuous objective function used in [5]. More critically, cardinality maximisation is inherently robust to 2D and 3D outliers, while avoiding the problems associated with trimming. The latter requires the user to specify the inlier fraction, which can rarely be known and is less intuitive to select than a geometrically meaningful inlier threshold. If the inlier fraction is over- or under-estimated, this approach may converge to the wrong pose, without a means to detect failure. Figure 2 demonstrates how the global optimum of a trimmed objective function, as used by [5, 49], may not occur at the true pose, a problem that is exacerbated when the inlier fraction is guessed incorrectly.

2. Related Work

A large body of work exists for solving the 2D–3D registration problem when correspondences are provided. When the correspondences are known perfectly, Perspective-n-Point (PnP) solvers [16, 26, 23, 18, 22] are able to estimate the pose of a camera given a set of noisy image points and their corresponding 3D points. When outliers are present in

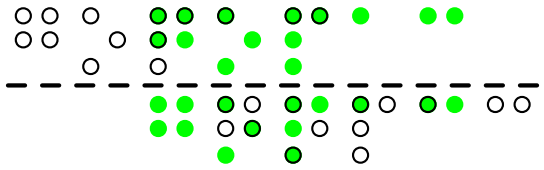


Figure 2. Two zero-error but incorrect 1D alignments of 2 point-sets with 8 trimmed 'outliers'. With noise, the global optimum of a trimmed objective function may not occur at the true pose, particularly if an incorrect trimming fraction is selected. The problem is exacerbated with higher dimensions and degrees of freedom.

the correspondence set, the RANSAC framework [13, 8] or robust global optimisation [27, 11, 1, 48, 12, 47] can be used to find the inlier set. Alternatively, outlier removal schemes can make the problem more tractable [46, 40, 50, 7]. Other methods develop sophisticated matching strategies to avoid outlier correspondences at the outset [30, 44, 45, 29]. However, these methods require some correct correspondences. For this reason, they are often only practical for 3D models that have been constructed using stereopsis or Structure-from-Motion (SfM). These models associate an image feature with each 3D point, facilitating inter-modality feature matching. Generic point-sets do not have this property; a point may lie anywhere on the underlying surfaces in a laser scan, not just where strong image gradients occur.

When correspondences are unknown, the problem becomes more challenging. For the 2D–2D case, problems such as correspondence-free rigid registration [3, 4], SfM [10, 33, 31] and relative camera pose [14] have been addressed. For the 2D–3D case, solution have been proposed for registering a collection of images [43] or multiple cameras [42] to a 3D point-set. The more general problem, however, is pose estimation from a single image. David *et al*. [9] proposed the SoftPOSIT algorithm, which alternates correspondence assignment with an iterative pose update algorithm. Moreno-Noguer *et al*. [35] proposed the BlindPnP algorithm, which represents the pose prior as a Gaussian mixture model from which a Kalman filter is initialised for matching. It outperformed SoftPOSIT when large amounts of clutter, occlusions and repetitive patterns were present. However, both are susceptible to local optima, require a pose prior and cannot guarantee global optimality.

Grimson [15] applied a RANSAC-like approach to the correspondence-free case, removing the need for a pose prior, but the method is not optimal and quickly becomes intractable as the number of points increase. In contrast, globally-optimal methods find a camera pose that is guaranteed to be an optimiser of an error function without requiring a pose prior, but tractability remains a challenge. A Branch-and-Bound (BB) [25] strategy may be applied in these cases, for which bounds need to be derived. For example, Breuel [4] used BB for 2D–2D registration problems, Hartley and Kahl [17] for optimal relative pose estimation by bounding the group of 3D rotations, Li and Hartley [28]

for rotation-only 3D–3D registration, Olsson *et al*. [41] for 3D–3D registration with known correspondences, Yang *et al*. [49] for full 3D–3D registration and Campbell and Petersson [6] for robust 3D–3D registration. While not optimal, Jurie [20] used an approach similar to BB for 2D–3D alignment with a linear approximation of perspective projection. Brown *et al*. [5] proposed a global and ϵ-suboptimal method using BB. It finds a camera pose whose trimmed geometric error, the sum of angular distances between the bearings and their rotationally-closest 3D points, is within ϵ of the global minimum. While not susceptible to local minima, it requires the inlier fraction to be specified, which can rarely be known in advance, in order to trim outliers.

Our work is the first globally-optimal inlier set cardinality maximisation solution to the simultaneous pose and correspondence problem. It is guaranteed to find the exact global optimum without requiring a pose prior and is robust to 2D and 3D outliers while avoiding the distortion of trimming. The rest of the paper is organised as follows: we introduce the problem formulation in Section 3, develop a parametrisation of the domain of 3D motions, a branching strategy and a derivation of the bounds in Section 4, propose an algorithm for globally-optimal pose and correspondence in Section 5 and evaluate its performance in Section 6.

3. Inlier Set Cardinality Maximisation

Let $\mathbf{p} \in \mathbb{R}^3$ be a 3D point and $\mathbf{f} \in \mathbb{R}^3$ be a bearing vector with unit norm, corresponding to a 2D point imaged by a calibrated camera. That is, $\mathbf{f} \propto \mathtt{K}^{-1}\hat{\mathbf{x}}$ where $\mathtt{K}$ is the matrix of intrinsic camera parameters and $\hat{\mathbf{x}}$ is the homogeneous image point. Given a set of points $\mathcal{P} = \{\mathbf{p}_j\}_{j=1}^{M}$ and bearing vectors $\mathcal{F} = \{\mathbf{f}_i\}_{i=1}^{N}$ and an inlier threshold θ, the objective is to find a rotation $\mathtt{R} \in SO(3)$ and translation $\mathbf{t} \in \mathbb{R}^3$ that maximises the cardinality ν of the inlier set $\mathcal{S}_I$

$$\nu^* = \max_{\mathtt{R},\,\mathbf{t}} |\mathcal{S}_I| \tag{1}$$

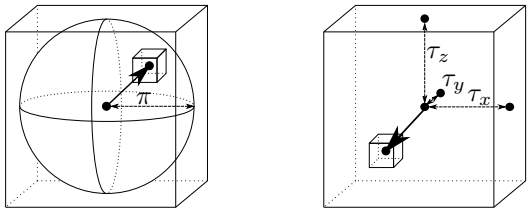
$$\mathcal{S}_I = \{\mathbf{f} \in \mathcal{F} \mid \exists \mathbf{p} \in \mathcal{P} : \angle(\mathbf{f}, \mathtt{R}(\mathbf{p} - \mathbf{t})) \leqslant \theta\} \tag{2}$$

where $\angle(\cdot,\cdot)$ denotes the angular distance between vectors. An equivalent formulation is given by

$$\nu^* = \max_{\mathtt{R},\,\mathbf{t}} f(\mathtt{R}, \mathbf{t}) \tag{3}$$

$$f(\mathtt{R}, \mathbf{t}) = \sum_{\mathbf{f} \in \mathcal{F}} \max_{\mathbf{p} \in \mathcal{P}} \mathbf{1}\big(\theta - \angle(\mathbf{f}, \mathtt{R}(\mathbf{p} - \mathbf{t}))\big) \tag{4}$$

where $\mathbf{1}(x) \triangleq \mathbf{1}_{\mathbb{R}_{\geq 0}}(x)$ is the indicator function that has the value 1 for all elements of the non-negative real numbers and the value 0 otherwise. The optimal transformation parameters $\mathtt{R}^*$ and $\mathbf{t}^*$ allow us to find all correspondences $(\mathbf{f}_i, \mathbf{p}_j)$ with respect to θ by identifying all pairs for which $\angle(\mathbf{f}_i, \mathtt{R}^*(\mathbf{p}_j - \mathbf{t}^*)) \leqslant \theta$. We maximise the cardinality of the set of bearing vector inliers, not the set of 3D point inliers, to avoid the degenerate case of all points sharing the same bearing vector inlier, which occurs when the camera is translated far away from the point-set.



(a) Rotation Domain Ω_r (b) Translation Domain Ω_t

Figure 3. Parametrisation of $SE(3)$. (a) The rotation space $SO(3)$ is parametrised by angle-axis 3-vectors in a solid radius-π ball. (b) The translation space $\mathbb{R}^3$ is parametrised by 3-vectors bounded by a cuboid with half-widths $[\tau_x, \tau_y, \tau_z]$. The domain is branched into sub-cuboids as shown using nested octree data structures.

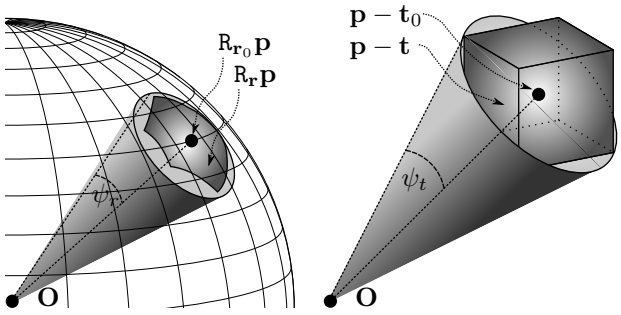
4. Branch-and-Bound

To solve the highly non-convex cardinality maximisation problem (1), the global optimisation technique of Branch-and-Bound (BB) [25] may be applied. To do so, a suitable means of parametrising and branching (partitioning) the function domain must be found, as well as an efficient way to calculate upper and lower bounds of the function for each branch which converge as the size of the branch tends to zero. While the bounds need to be computationally efficient to calculate, the time and memory efficiency of the algorithm also depends on how tight the bounds are, since tighter bounds reduce the search space quicker by allowing suboptimal branches to be pruned.

4.1. Parametrising and Branching the Domain

To find a globally-optimal solution, the cardinality of the inlier set $\mathcal{S}_I$ must be maximised over the domain of 3D motions, that is, the group $SE(3) = SO(3) \times \mathbb{R}^3$. However, the space of these transformations is unbounded, therefore we restrict the space of translations to be within the bounded set Ω_t in order to use BB. For a suitably large Ω_t, it is reasonable to assume that the camera centre lies within Ω_t. That is, we can assume that the camera is less than a finite distance from the 3D points. The domains are shown in Figure 3.

Rotation space $SO(3)$ is minimally parametrised with angle-axis 3-vectors $\mathbf{r}$ with rotation angle $\|\mathbf{r}\|$ and rotation axis $\mathbf{r}/\|\mathbf{r}\|$. The notation $\mathtt{R}_\mathbf{r} \in SO(3)$ is used to denote the rotation matrix obtained from the matrix exponential map of the skew-symmetric matrix $[\mathbf{r}]_\times$ induced by $\mathbf{r}$. The Rodrigues' rotation formula can be used to efficiently calculate this mapping. Using this parametrisation, the space of all 3D rotations can be represented as a solid ball of radius π in $\mathbb{R}^3$. The mapping is one-to-one on the interior of the π-ball and two-to-one on the surface. For ease of manipulation, we use the 3D cube circumscribing the π-ball as the rotation domain Ω_r [28]. Translation space $\mathbb{R}^3$ is parametrised with 3-vectors in a bounded domain chosen as the cuboid Ω_t containing the bounding box of $\mathcal{P}$. If the camera is known to be inside the 3D scene, Ω_t can be set to the bounding box, otherwise it is set to an expansion of the bounding box.



(a) Rotation Uncertainty Angle (b) Translation Uncertainty Angle

Figure 4. Uncertainty angles induced by rotation and translation sub-cubes. (a) Rotation uncertainty angle ψ_r for $\mathcal{C}_r$. The optimal rotation of $\mathbf{p}$ may be anywhere within the umbrella-shaped region, which is entirely contained by the cone defined by $\mathtt{R}_{\mathbf{r}_0}\mathbf{p}$ and ψ_r. (b) Translation uncertainty angle ψ_t for $\mathcal{C}_t$. The optimal translation of $\mathbf{p}$ may be anywhere within the cuboidal region, which is entirely contained by the cone defined by $\mathbf{p}-\mathbf{t}_0$ and ψ_t.

During BB, the domain is branched into sub-cuboids using nested octree data structures. They are defined as

$$\mathcal{C}(\mathbf{c},\boldsymbol{\delta}) = \{\mathbf{x}\in\mathbb{R}^3 \mid \mathbf{e}_i^\intercal(\mathbf{x}-\mathbf{c}) \in [-\delta_i,\delta_i], i=1,2,3\} \quad (5)$$

where $\mathbf{e}_i$ is the i^{th} standard basis vector. To simplify the notation, we use $\mathcal{C}_r = \mathcal{C}(\mathbf{r}_0,\boldsymbol{\delta}_r)$ and $\mathcal{C}_t = \mathcal{C}(\mathbf{t}_0,\boldsymbol{\delta}_t)$.

The uncertainty angle induced by a rotation and translation sub-cuboid on a point $\mathbf{p}$ is shown in Figure 4. The transformed point may lie anywhere within an uncertainty cone, with aperture angle equal to the sum of the rotation and translation uncertainty angles.

4.2. Bounding the Branches

The success of a BB algorithm is predicated on the quality of its bounds. For inlier set maximisation, the objective function (4) needs to be bounded within a transformation domain. Some preparatory material is now presented.

To bound the uncertainty angle due to rotation, Lemmas 1 and 2 from [17] are used. For reference, the relevant parts are merged into Lemma 1, as in [49]. The lemma indicates that the angle between two rotated vectors is less than or equal to the Euclidean distance between their rotations' angle-axis representations in $\mathbb{R}^3$.

Lemma 1. *For an arbitrary vector $\mathbf{p}$ and two rotations, represented as $\mathtt{R}_{\mathbf{r}_1}$ and $\mathtt{R}_{\mathbf{r}_2}$ in matrix form and $\mathbf{r}_1$ and $\mathbf{r}_2$ in angle-axis form,*

$$\angle(\mathtt{R}_{\mathbf{r}_1}\mathbf{p},\, \mathtt{R}_{\mathbf{r}_2}\mathbf{p}) \leqslant \|\mathbf{r}_1-\mathbf{r}_2\|. \quad (6)$$

From this, the maximum angle between a vector $\mathbf{p}$ rotated by $\mathbf{r}_0$ and $\mathbf{p}$ rotated by $\mathbf{r}\in\mathcal{C}_r$ can be found as follows.

Lemma 2. *(Weak rotation uncertainty angle) Given a 3D point $\mathbf{p}$ and a rotation cube $\mathcal{C}_r$ of half side-length δ_r centred at $\mathbf{r}_0$, then $\forall\mathbf{r}\in\mathcal{C}_r$,*

$$\angle(\mathtt{R}_{\mathbf{r}}\mathbf{p},\mathtt{R}_{\mathbf{r}_0}\mathbf{p}) \leqslant \min(\sqrt{3}\delta_r,\pi) \triangleq \psi_r^w(\mathcal{C}_r). \quad (7)$$

Proof. Inequality (7) can be derived as follows:

$$\angle(\mathtt{R}_{\mathbf{r}}\mathbf{p},\mathtt{R}_{\mathbf{r}_0}\mathbf{p}) \leqslant \min(\|\mathbf{r}-\mathbf{r}_0\|,\pi) \quad (8)$$
$$\leqslant \min(\sqrt{3}\delta_r,\pi) \quad (9)$$

where (8) follows from Lemma 1 and the maximum possible angle and (9) follows from $\max\|\mathbf{r}-\mathbf{r}_0\| = \sqrt{3}\delta_r$ (the half space diagonal of the rotation cube) for $\mathbf{r}\in\mathcal{C}_r$. □

However, a tighter bound can be found by observing that a point rotated about an axis parallel to the point is not displaced. To exploit this, we maximise the angle $\angle(\mathtt{R}_{\mathbf{r}}\mathbf{p},\mathtt{R}_{\mathbf{r}_0}\mathbf{p})$ over the surface $\mathcal{S}_r$ of the cube $\mathcal{C}_r$.

Lemma 3. *(Rotation uncertainty angle) Given a 3D point $\mathbf{p}$ and a rotation cube $\mathcal{C}_r$ centred at $\mathbf{r}_0$ with surface $\mathcal{S}_r$, then $\forall\mathbf{r}\in\mathcal{C}_r$,*

$$\angle(\mathtt{R}_{\mathbf{r}}\mathbf{p},\mathtt{R}_{\mathbf{r}_0}\mathbf{p}) \leqslant \min(\max_{\mathbf{r}\in\mathcal{S}_r}\angle(\mathtt{R}_{\mathbf{r}}\mathbf{p},\mathtt{R}_{\mathbf{r}_0}\mathbf{p}),\pi) \triangleq \psi_r(\mathbf{p},\mathcal{C}_r). \quad (10)$$

Proof. Inequality (10) can be derived as follows:

$$\angle(\mathtt{R}_{\mathbf{r}}\mathbf{p},\mathtt{R}_{\mathbf{r}_0}\mathbf{p}) \leqslant \min(\max_{\mathbf{r}\in\mathcal{C}_r}\angle(\mathtt{R}_{\mathbf{r}}\mathbf{p},\mathtt{R}_{\mathbf{r}_0}\mathbf{p}),\pi) \quad (11)$$
$$= \min(\max_{\mathbf{r}\in\mathcal{S}_r}\angle(\mathtt{R}_{\mathbf{r}}\mathbf{p},\mathtt{R}_{\mathbf{r}_0}\mathbf{p}),\pi) \quad (12)$$

where (12) is a consequence of the order-preserving mapping, with respect to the radial angle, from the convex cube of angle-axis vectors to the spherical surface patch (see Figure 4a), since the mapping is obtained by projecting from the centre of the sphere to the surface of the sphere. See the appendix for further details. □

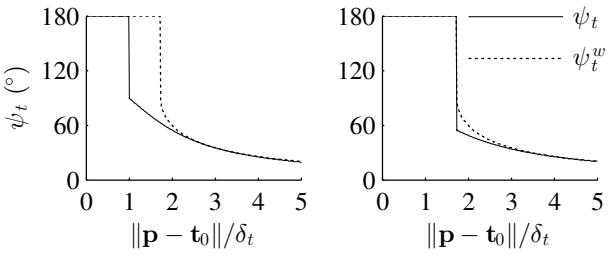
The uncertainty angle due to translation can be bounded by observing that the translated points form a cube (Figure 4b). When the cube does not contain the origin, the angle can be found by maximising over the cube vertices.

Lemma 4. *(Translation uncertainty angle) Given a 3D point $\mathbf{p}$ and a translation cube $\mathcal{C}_t$ centred at $\mathbf{t}_0$ with vertices $\mathcal{V}_t$, then $\forall\mathbf{t}\in\mathcal{C}_t$,*

$$\angle(\mathbf{p}-\mathbf{t},\mathbf{p}-\mathbf{t}_0) \leqslant \begin{cases}\max\limits_{\mathbf{t}\in\mathcal{V}_t}\angle(\mathbf{p}-\mathbf{t},\mathbf{p}-\mathbf{t}_0) & \text{if } \mathbf{p}\notin\mathcal{C}_t\\ \pi & \text{else}\end{cases}$$
$$\triangleq \psi_t(\mathbf{p},\mathcal{C}_t). \quad (13)$$

Proof. Observe that for $\mathbf{p}\in\mathcal{C}_t$, the cube containing all translated points $\mathbf{p}-\mathbf{t}$ also contains the origin. Therefore $\mathbf{p}-\mathbf{t}$ can be proportional to $-(\mathbf{p}-\mathbf{t}_0)$ and thus the maximum angle is π. For $\mathbf{p}\notin\mathcal{C}_t$,

$$\angle(\mathbf{p}-\mathbf{t},\mathbf{p}-\mathbf{t}_0) \leqslant \max_{\mathbf{t}\in\mathcal{C}_t}\angle(\mathbf{p}-\mathbf{t},\mathbf{p}-\mathbf{t}_0) \quad (14)$$
$$= \max_{\mathbf{t}\in\mathcal{V}_t}\angle(\mathbf{p}-\mathbf{t},\mathbf{p}-\mathbf{t}_0) \quad (15)$$



(a) Ray through face centre (b) Ray through vertex

Figure 5. Comparison of translation bounds when the cube centre lies along a ray from the origin towards (a) any face centre and (b) any vertex. Our bound ψ_t is tighter across the entire domain.

where (15) follows from the convexity of the angle function in this domain. The maximum of a convex function over a convex set must occur at one of its extreme points (the vertices). Geometrically, the cube $\mathbf{p}-\mathbf{t}$ projects to a spherical hexagon on the unit sphere. The maximum geodesic from a point in the hexagon to any other is to a vertex. □

To avoid the non-physical case where a 3D point is located within a very small value ζ of the camera centre we restrict the translation domain such that $\Omega_t' = \Omega_t \cap \{\mathbf{t} \in \mathbb{R}^3 \mid \|\mathbf{p}-\mathbf{t}\| \geqslant \zeta, \forall \mathbf{p} \in \mathcal{P}\}$.

The translation bound from [5] encloses a translation cube with a sphere of radius $\rho_t = \sqrt{3}\delta_t$ and is given by

$$\psi_t^w(\mathbf{p}, \mathcal{C}_t) \triangleq \begin{cases} \arcsin\left(\frac{\rho_t}{\|\mathbf{p}-\mathbf{t}_0\|}\right) & \text{if } \rho_t \leqslant \|\mathbf{p}-\mathbf{t}_0\| \\ \pi & \text{else.} \end{cases} \quad (16)$$

Our bound is tighter with a maximum difference of 117° for cubes and greater for cuboids. Figure 5 compares both translation bounds across a range of values.

The preceding lemmas are used to bound the objective function (4) within a transformation domain $\mathcal{C}_r \times \mathcal{C}_t$. For brevity, we use the notation $\mathbf{p}_\mathbf{t}^\mathbf{r} \triangleq \mathtt{R}_\mathbf{r}(\mathbf{p}-\mathbf{t})$, $\mathbf{p}_\mathbf{t} \triangleq \mathbf{p}-\mathbf{t}$ and $\mathbf{f}_\mathbf{r} \triangleq (\mathtt{R}_\mathbf{r})^{-1}\mathbf{f}$.

Theorem 1. *(Lower bound) For the domain $\mathcal{C}_r \times \mathcal{C}_t$ centred at $(\mathbf{r}_0, \mathbf{t}_0)$, the lower bound of the inlier set cardinality can be chosen as*

$$\underline{f}(\mathtt{R}_\mathbf{r}, \mathbf{t}) \triangleq f(\mathtt{R}_{\mathbf{r}_0}, \mathbf{t}_0). \quad (17)$$

Proof. The validity of the lower bound follows from

$$\max_{\mathbf{r},\mathbf{t}} f(\mathtt{R}_\mathbf{r}, \mathbf{t}) \geqslant f(\mathtt{R}_{\mathbf{r}_0}, \mathbf{t}_0). \quad (18)$$

That is, the function value at a specific point within the domain is less than or equal to the maximum. □

Theorem 2. *(Upper bound) For the domain $\mathcal{C}_r \times \mathcal{C}_t$ centred at $(\mathbf{r}_0, \mathbf{t}_0)$, the upper bound of the inlier set cardinality can be chosen as*

$$\bar{f}(\mathtt{R}_\mathbf{r}, \mathbf{t}) \triangleq \sum_{\mathbf{f}\in\mathcal{F}} \max_{\mathbf{p}\in\mathcal{P}} \mathbf{1}\big(\theta - \angle(\mathbf{f}, \mathbf{p}_{\mathbf{t}_0}^{\mathbf{r}_0}) + \psi_r(\mathbf{f}, \mathcal{C}_r) + \psi_t(\mathbf{p}, \mathcal{C}_t)\big). \quad (19)$$

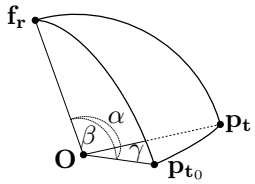


Figure 6. The triangle inequality in spherical geometry, given by $\beta \leqslant \alpha + \gamma$ or $\angle(\mathbf{f_r}, \mathbf{p_{t_0}}) \leqslant \angle(\mathbf{f_r}, \mathbf{p_t}) + \angle(\mathbf{p_t}, \mathbf{p_{t_0}})$. The transformed points have been normalised to lie on the unit sphere.

Proof. Observe that $\forall(\mathbf{r}, \mathbf{t}) \in (\mathcal{C}_r \times \mathcal{C}_t)$,

$$\angle(\mathbf{f}, \mathbf{p}_\mathbf{t}^\mathbf{r}) \geqslant \angle(\mathbf{f}, \mathbf{p}_{\mathbf{t}_0}^{\mathbf{r}_0}) - \angle(\mathbf{f}_\mathbf{r}, \mathbf{f}_{\mathbf{r}_0}) - \angle(\mathbf{p_t}, \mathbf{p}_{\mathbf{t}_0}) \quad (20)$$

$$\geqslant \angle(\mathbf{f}, \mathbf{p}_{\mathbf{t}_0}^{\mathbf{r}_0}) - \psi_r(\mathbf{f}, \mathcal{C}_r) - \psi_t(\mathbf{p}, \mathcal{C}_t) \quad (21)$$

where (20) follows from the triangle inequality in spherical geometry (see Figure 6) and (21) follows from Lemmas 3 and 4. Substituting (21) into (4) completes the proof. □

By inspecting the translation component of Theorem 2, a tighter upper bound may be found by removing one of the two applications of the triangle inequality. A similar approach cannot be taken for the rotation component since $\mathtt{R}_\mathbf{r}\mathbf{p}$ is a complex surface due to the nonlinear conversion from angle-axis to rotation matrix representations. To reduce computation, it is only necessary to evaluate this tighter bound when $\angle(\mathbf{f}, \mathbf{p}_{\mathbf{t}_0}^{\mathbf{r}_0}) \leqslant \theta + \psi_r(\mathbf{f}, \mathcal{C}_r) + \psi_t(\mathbf{p}, \mathcal{C}_t)$, since otherwise the point is definitely an outlier and does not need to be investigated further.

Theorem 3. *(Tighter upper bound) For the domain $\mathcal{C}_r \times \mathcal{C}_t$ centred at $(\mathbf{r}_0, \mathbf{t}_0)$, the upper bound of the inlier set cardinality can be chosen as*

$$\bar{f}(\mathtt{R}_\mathbf{r}, \mathbf{t}) \triangleq \sum_{\mathbf{f}\in\mathcal{F}} \max_{\mathbf{p}\in\mathcal{P}} \Gamma(\mathbf{f}, \mathbf{p}) \quad (22)$$

$$\Gamma(\mathbf{f}, \mathbf{p}) = \max_{\mathbf{t}\in\mathcal{C}_t} \mathbf{1}\big(\theta - \angle(\mathbf{f}, \mathbf{p}_\mathbf{t}^{\mathbf{r}_0}) + \psi_r(\mathbf{f}, \mathcal{C}_r)\big) \quad (23)$$

Proof. Observe that $\forall(\mathbf{r}, \mathbf{t}) \in (\mathcal{C}_r \times \mathcal{C}_t)$,

$$\mathbf{1}\big(\theta - \angle(\mathbf{f}, \mathbf{p}_\mathbf{t}^\mathbf{r})\big) \leqslant \mathbf{1}\big(\theta - \angle(\mathbf{f}, \mathbf{p}_\mathbf{t}^{\mathbf{r}_0}) + \angle(\mathbf{f}_\mathbf{r}, \mathbf{f}_{\mathbf{r}_0})\big) \quad (24)$$

$$\leqslant \max_{\mathbf{t}\in\mathcal{C}_t} \mathbf{1}\big(\theta - \angle(\mathbf{f}, \mathbf{p}_\mathbf{t}^{\mathbf{r}_0}) + \psi_r(\mathbf{f}, \mathcal{C}_r)\big) \quad (25)$$

where (24) follows from the triangle inequality in spherical geometry and (25) follows from Lemma 3 and maximising over $\mathbf{t}$. Substituting (25) into (4) completes the proof. □

Γ may be evaluated by observing that the minimum angle between a ray $\mathbf{f}$ and a cube $\mathbf{p}_\mathbf{t}^{\mathbf{r}_0}$ is zero if the ray passes through the cube and is otherwise the angle between the ray and the point on the skeleton of the cube (vertices and edges) with least angular displacement from $\mathbf{f}$. Thus, for the translation domain $\mathcal{C}_t$ with skeleton $\mathcal{S}k_t$,

$$\Gamma = \begin{cases} \max_{\mathbf{t}\in\mathcal{S}k_t} \mathbf{1}\big(\theta - \angle(\mathbf{f}, \mathbf{p}_\mathbf{t}^{\mathbf{r}_0}) + \psi_r\big) & \text{if } \angle(\mathbf{f}, \mathbf{p}_{\mathbf{t}_0}^{\mathbf{r}_0}) > \psi_t \\ 1 & \text{else.} \end{cases} \quad (26)$$

5. The GOPAC Algorithm

The Globally-Optimal Pose And Correspondences (GOPAC) algorithm for a calibrated camera is outlined in Algorithms 1 and 2. As in [49], we employ a nested branch-and-bound structure for computational efficiency. In the outer breadth-first BB search, upper and lower bounds are found for each translation cuboid $\mathcal{C}_t \in \Omega_t$ by running an inner BB search over rotation space $SO(3)$ (denoted RBB). The upper bound $\bar{\nu} \triangleq \bar{\nu}_t$ (19) of $\mathcal{C}_t$ is found by running RBB until convergence with the following bounds

$$\underline{\nu}_r \triangleq \sum_{\mathbf{f}\in\mathcal{F}} \max_{\mathbf{p}\in\mathcal{P}} \mathbf{1}\big(\theta - \angle(\mathbf{f}, \mathbf{p}_{\mathbf{t}_0}^{\mathbf{r}_0}) + \psi_t(\mathbf{p})\big) \tag{27}$$

$$\bar{\nu}_r \triangleq \sum_{\mathbf{f}\in\mathcal{F}} \max_{\mathbf{p}\in\mathcal{P}} \mathbf{1}\big(\theta - \angle(\mathbf{f}, \mathbf{p}_{\mathbf{t}_0}^{\mathbf{r}_0}) + \psi_t(\mathbf{p}) + \psi_r(\mathbf{f})\big). \tag{28}$$

The tighter upper bound (22) instead uses

$$\underline{\nu}_r \triangleq \sum_{\mathbf{f}\in\mathcal{F}} \max_{\mathbf{p}\in\mathcal{P},\mathbf{t}\in\mathcal{C}_t} \mathbf{1}\big(\theta - \angle(\mathbf{f}, \mathbf{p}_{\mathbf{t}}^{\mathbf{r}_0})\big) \tag{29}$$

$$\bar{\nu}_r \triangleq \sum_{\mathbf{f}\in\mathcal{F}} \max_{\mathbf{p}\in\mathcal{P},\mathbf{t}\in\mathcal{C}_t} \mathbf{1}\big(\theta - \angle(\mathbf{f}, \mathbf{p}_{\mathbf{t}}^{\mathbf{r}_0}) + \psi_r(\mathbf{f})\big). \tag{30}$$

The lower bound $\underline{\nu} \triangleq \underline{\nu}_t$ (17) is found by running RBB using bounds (27) and (28) with ψ_t set to zero.

The nested structure has better memory and computational efficiency than directly branching over 6D transformation space, since it maintains a queue for each 3D sub-problem, rather than one for the entire 6D problem. This requires significantly fewer simultaneously enqueued sub-cubes. Moreover, with rotation search nested inside translation search, ψ_t only has to be calculated once per translation $\mathbf{t}$, not once per pose $(\mathbf{r}, \mathbf{t})$, and $\mathcal{F}$ can be rotated (by $\mathtt{R}^{-1}$) instead of $\mathcal{P}$ which typically has more elements. This makes it possible to precompute the rotated bearing vectors and rotation bounds for the top five levels of the rotation octree to reduce the amount of computation required in the inner BB subroutine.

Line 9 of Algorithm 1 shows how local optimisation is incorporated to refine the camera pose, in a similar manner to [49, 5]. Whenever the BB algorithm finds a sub-cube pair $(\mathcal{C}_r, \mathcal{C}_t)$ with a greater lower bound $\underline{\nu}$ than half the best-so-far cardinality ν^*, the PnP problem is solved, with correspondences given by the inlier pairs at the pose $(\mathbf{r}_0, \mathbf{t}_0)$. We use nonlinear optimisation [21], minimising the sum of angular distances between corresponding bearing vectors and points, and update ν^* if a larger ν is found. In this way, BB and PnP collaborate, with PnP finding the best pose given correspondences and BB guiding the search for correspondences. PnP accelerates convergence since the faster ν^* is increased, the sooner sub-cubes (with $\bar{\nu} \leqslant \nu^*$) can be culled (Alg. 1 Line 11). SoftPOSIT [9] is also applied at this stage to help jump out of local minima.

Algorithm 1 GOPAC: a branch-and-bound algorithm for globally-optimal camera pose & correspondence estimation

Input: bearing vector set $\mathcal{F}$, point set $\mathcal{P}$, inlier threshold θ, initial domains Ω_r and Ω_t

Output: optimal number of inliers ν^*, camera pose $(\mathbf{r}^*, \mathbf{t}^*)$ and 2D–3D correspondences

1: $\nu^* \leftarrow 0$
2: Add translation domain Ω_t to priority queue Q_t
3: **loop**
4: Update greatest upper bound $\bar{\nu}_t$ from Q_t
5: Get cuboid $\mathcal{C}_t$ with greatest width δ_{tx} from Q_t
6: **if** $\nu^* \geqslant \bar{\nu}_t$ **then** terminate
7: **for all** sub-cuboids $\mathcal{C}_{ti} \in \mathcal{C}_t$ **do**
8: $(\underline{\nu}_{ti}, \mathbf{r}) \leftarrow \mathrm{RBB}(\nu^*, \mathbf{t}_{0i}, \psi_t = 0)$
9: **if** $\nu^* < 2\underline{\nu}_{ti}$ **then** $(\nu^*, \mathbf{r}^*, \mathbf{t}^*) \leftarrow \mathrm{P}n\mathrm{P}(\mathbf{r}, \mathbf{t}_{0i})$
10: $\bar{\nu}_{ti} \leftarrow \mathrm{RBB}(\nu^*, \mathbf{t}_{0i}, \psi_t)$
11: **if** $\nu^* < \bar{\nu}_{ti}$ **then** add $\mathcal{C}_{ti}$ to queue Q_t

Algorithm 2 RBB: a rotation search subroutine for GOPAC

Input: bearing vector set $\mathcal{F}$, point set $\mathcal{P}$, inlier threshold θ, initial domain Ω_r, best-so-far cardinality ν^*, translation $\mathbf{t}_0$, translation uncertainty ψ_t

Output: optimal number of inliers ν_r^* and rotation $\mathtt{R}^*$

1: $\nu_r^* \leftarrow \nu^*$
2: Add rotation domain Ω_r to priority queue Q_r
3: **loop**
4: Read cube $\mathcal{C}_r$ with greatest upper bound $\bar{\nu}_r$ from Q_r
5: **if** $\nu_r^* \geqslant \bar{\nu}_r$ **then** terminate
6: **for all** sub-cubes $\mathcal{C}_{ri} \in \mathcal{C}_r$ **do**
7: Calculate $\underline{\nu}_{ri}$ by (27) or (29)
8: **if** $\nu_r^* < \underline{\nu}_{ri}$ **then** $\nu_r^* \leftarrow \underline{\nu}_{ri}, \mathbf{r}^* \leftarrow \mathbf{r}_0$
9: Calculate $\bar{\nu}_{ri}$ by (28) or (30)
10: **if** $\nu_r^* < \bar{\nu}_{ri}$ **then** add $\mathcal{C}_{ri}$ to queue Q_r

As just observed, a large ν^* reduces runtime. Therefore, if the user knows a lower bound on the number of 2D inliers, ν^* can be initialised to this value. However, this is rarely known. Instead, our algorithm implements an optional guess-and-verify approach, without loss of optimality or objective function distortion, which provides especial benefit when 2D outliers are rare: set $\nu^* = n$; run GOPAC; stop if an optimality guarantee is found, otherwise $n \leftarrow \max(n - s, 0)$ and repeat. We initialise $n = N - 1$ and $s = \lceil 0.1N \rceil$.

We also provide a multi-threaded implementation, where the initial translation domain is divided into sub-domains and GOPAC is run for each in separate CPU threads. The algorithm returns the largest ν^* and the associated pose and correspondences. While not supplied, a massively parallel implementation on a GPU is very feasible. Further algorithmic details are provided in the appendix.

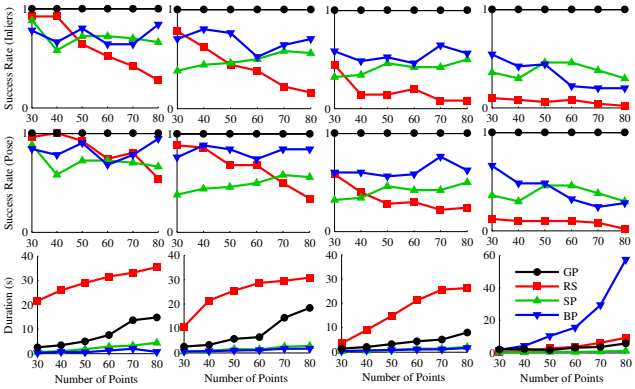
6. Results

The GOPAC algorithm was evaluated with respect to the baseline RANSAC [13], SoftPOSIT [9] and BlindPnP [35] algorithms, denoted GP, RS, SP and BP respectively, with synthetic and real data. The RANSAC approach uses the OpenGV framework [21] and the P3P algorithm [23] with randomly-sampled correspondences. Since SoftPOSIT and BlindPnP require pose priors to function, we use a torus prior in the synthetic experiments. In general, the space of camera poses is much larger than the restrictive torus prior and a good prior can rarely be known in advance. Except where otherwise specified, the inlier threshold θ was set to 1°, the rotation and translation bounds (10) and (13) were used, SoftPOSIT and nonlinear PnP refinement were applied and multithreading was not used. It is crucial to observe that finding the global optimum does not necessarily imply finding the ground-truth transformation. There may be multiple global optima, particularly in the case of symmetries, and noise may create false optima.

6.1. Synthetic Data Experiments

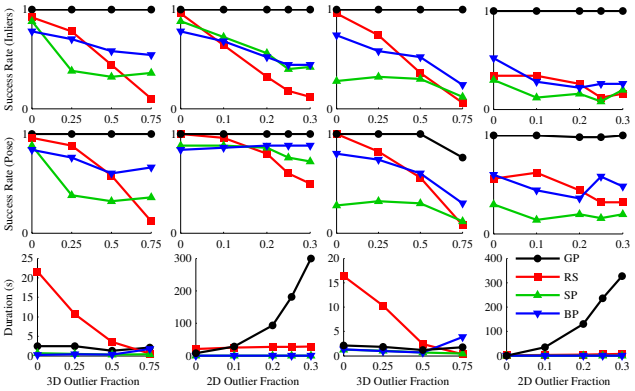
To evaluate our algorithm in a setting where true priors can be applied, we performed 50 independent Monte Carlo simulations per parameter setting, using the framework of [35]: M random 3D points were generated from $[-1, 1]^3$; a fraction ω_{3D} of the 3D points were randomly selected as outliers to model occlusion; the inliers were projected to a virtual image; normal noise was added with $\sigma = 2$ pixels; and random points were added to the image such that a fraction ω_{2D} of the 2D points were outliers. To facilitate fair comparison with SoftPOSIT and BlindPnP, we use a pose prior for these experiments. The torus prior constrains the camera centre to a torus around the point-set with the optical axis directed towards the model, as in [35]. BlindPnP represents the poses with a 20 component Gaussian mixture model, the means of which are used to initialise SoftPOSIT, as in [35]. GOPAC is given a set of translation cubes which approximate the torus and is not given the rotation priors.

The results are shown in Figures 7 and 8a. We repeated the experiments for the repetitive CAD structure shown in Figure 9a, with results shown in Figure 8b. Two success rates are reported: the fraction of trials where the true maximum number of inliers was found and the fraction where the correct pose was found, where the angle between the output rotation and the ground truth rotation is less than 0.1 radians and the camera centre error $\|\mathbf{t} - \mathbf{t}_{GT}\|/\|\mathbf{t}_{GT}\|$ relative to the ground truth $\mathbf{t}_{GT}$ is less than 0.1, as in [35]. The 2D and 3D outlier fractions were fixed to 0 when not being varied and multithreading was used in the 2D outlier experiments. GOPAC outperforms the other methods, reliably finding the global optimum while still being relatively efficient, particularly when the fraction of 2D outliers is low. For the repetitive CAD structure, while GOPAC finds the globally optimal number of inliers in all cases, the pose is occasionally incorrect when 75% of the 3D points are occluded, due to the highly symmetric nature of the model.



(a) $\omega_{3D} = 0$ (b) $\omega_{3D} = 0.25$ (c) $\omega_{3D} = 0.5$ (d) $\omega_{3D} = 0.75$

Figure 7. Mean success rates and median runtimes with respect to the number of random 3D points and the 3D outlier fraction, for 50 Monte Carlo simulations per parameter value with the torus prior.

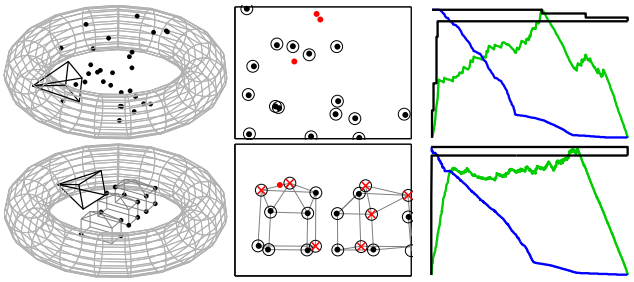


(a) Random Points $M = 30$ (b) CAD Structure $M = 27$

Figure 8. Mean success rates and median runtimes with respect to the 3D and 2D outlier fractions for the random points and CAD structure datasets, for 50 Monte Carlo simulations per parameter value with the torus prior.

The evolution of the global lower and upper bounds is shown in Figure 9c: BB and PnP collaborate to increase the lower bound with BB guiding the search into better convergence basins and PnP refining the bound by jumping to the nearest local maximum (the staircase pattern). The majority of the time is spent decreasing the upper bound, indicating it will often find the global optimum when terminated early.

To show the improvement attributable to the tighter upper bounds derived, we measured the runtime of the algorithm with 10 random 3D points and 50% 2D outliers using different upper bounds, shown in Figure 10. The weak sphere-based bounding functions in (7) and (16) are denoted ψ_r^w and ψ_t^w respectively, the tighter cuboid-based bounding functions in (10) and (13) are denoted ψ_r and ψ_t respectively and the bounding function from (22) is denoted Γ. Further results are provided in the appendix.



(a) 3D Models (b) 2D Alignment (c) Bound Evolution

Figure 9. Sample 2D and 3D results for two trials using the random points and repetitive CAD model datasets. (a) 3D models, true and GOPAC-estimated camera fulcra (completely overlapping) and toroidal pose priors. Only non-occluded 3D points are shown. (b) True projections of non-occluded 3D points are shown as black dots, 2D outliers as red dots, GOPAC projections as black circles and GOPAC-classified 3D outliers as red crosses. (c) Evolution over time of the upper and lower bounds (black), remaining translation volume (blue) and translation queue size (green) as a fraction of their maximum values. Best viewed in colour.

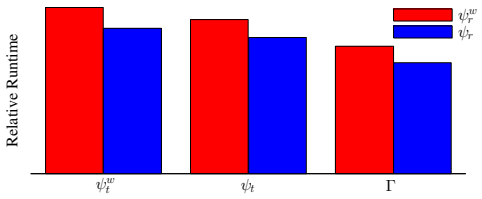


Figure 10. Comparison of the different upper bound functions. Runtime is plotted relative to the maximum (leftmost) value. The weakest upper bound is 50% slower than the tightest upper bound.

6.2. Real Data Experiments

To evaluate the algorithm on real data, we use the DATA61/2D3D (formerly NICTA) dataset [37], a large and repetitive multi-modal outdoor dataset. Finding the pose of a camera within a large laser-scanned point-set without a good initialisation represents an unsolved problem in computer vision, which this work makes progress towards solving. For each image, we obtain the ground truth camera pose from the provided 2D–3D correspondences using EPnP [26] followed by nonlinear PnP [21]. Extracting points from a laser scan that correspond to known pixels in an image is itself a challenging unsolved problem for 2D–3D registration pipelines. Due to the robust and optimal nature of GOPAC, we can relax this problem to isolating regions of the point-set that appear in the image and vice versa, from which putative correspondences may be drawn. We used semantic segmentations of the images and point-set to select regions that were potentially observable in both modalities, in this case the 'building' class. We then used grid downsampling and k-means clustering on the class pixels and points independently to reduce them to a manageable size and converted the pixels to bearing vectors. While we do not know the correspondences in advance, each bearing vector has a good chance of having a 3D point as an inlier. In this way, we constructed a dataset consisting of a 3D point-set with 88 points, a set of 11 images containing 30 2D features and a set of ground truth camera poses. For this experiment, we used an inlier threshold of $\theta = 2°$, multithreading and a 2D outlier fraction guess of $\omega_{2D} = 0.25$. The translation domain was $50 \times 5 \times 5$m, covering two lanes of the road, making use of the knowledge that the camera was mounted on a survey vehicle. SoftPOSIT and BlindPnP failed to find the correct camera pose for every image in this dataset, even when supplied the ground truth pose as a prior, due to the weak ground truth correspondences and an inability to handle 3D points behind the camera. Moreover, they do not natively support panoramic imagery and required an artificially restricted field of view to function.

Table 1. Camera pose results for the DATA61/2D3D dataset. The median translation error, rotation error and runtime and the mean inlier recall and success rates are reported. $\lfloor$GP$\rfloor$ denotes truncated GOPAC, where search is terminated after 30s, with no optimality guarantee. RS$_K$ denotes RANSAC with K million iterations.

Method	GP	$\lfloor$GP$\rfloor$	RS$_{20}$	RS$_{280}$
Translation Error (m)	**2.30**	3.10	20.3	28.5
Rotation Error (°)	**2.08**	3.04	178	179
Recall (Inliers)	**1.00**	0.97	0.75	0.81
Success Rate (Inliers)	**1.00**	0.45	0.00	0.00
Success Rate (Pose)	**0.82**	0.64	0.09	0.09
Runtime (s)	477	34	34	471

Qualitative results for the GOPAC and RANSAC algorithms are shown in Figure 1 and quantitative results in Table 1. GOPAC finds the optimal number of inliers for all frames and the correct camera pose for the majority of frames, despite the weakness of the 2D/3D point extraction process, surpassing the other methods. The failure modes for GOPAC were 180° rotation flips, due to ambiguities arising from the low angular separation of points in the vertical direction. The difficulty of this ill-posed problem is illustrated by the performance of truncated GOPAC, which was not able to find all optima even after running for 30s, motivating the necessity for globally-optimal guided search.

7. Conclusion

In this paper, we have introduced a robust and globally-optimal solution to the simultaneous camera pose and correspondence problem using inlier set cardinality maximisation. The method applies the branch-and-bound paradigm to guarantee optimality regardless of initialisation and uses local optimisation to accelerate convergence. The pivotal contribution is the derivation of the function bounds using the geometry of $SE(3)$. The algorithm outperformed other local and global methods on challenging synthetic and real datasets, finding the global optimum reliably. Further investigation is warranted to develop a complete 2D–3D pipeline, from segmentation and clustering to alignment.

References

[1] E. Ask, O. Enqvist, and F. Kahl. Optimal geometric fitting under the truncated L_2-norm. In *Proc. 2013 Conf. Comput. Vision Pattern Recognition*, pages 1722–1729. IEEE, 2013. 2

[2] M. Aubry, D. Maturana, A. A. Efros, B. C. Russell, and J. Sivic. Seeing 3D chairs: exemplar part-based 2D-3D alignment using a large dataset of CAD models. In *Proc. 2014 Conf. Comput. Vision Pattern Recognition*, pages 3762–3769, 2014. 1

[3] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, 1992. 2

[4] T. M. Breuel. Implementation techniques for geometric branch-and-bound matching methods. *Computer Vision and Image Understanding*, 90(3):258–294, 2003. 2

[5] M. Brown, D. Windridge, and J.-Y. Guillemaut. Globally optimal 2D-3D registration from points or lines without correspondences. In *Proc. 2015 Int. Conf. Comput. Vision*, pages 2111–2119, 2015. 2, 3, 5, 6

[6] D. Campbell and L. Petersson. GOGMA: Globally-Optimal Gaussian Mixture Alignment. In *Proc. 2016 Conf. Comput. Vision Pattern Recognition*, pages 5685–5694. IEEE, June 2016. 3

[7] T.-J. Chin, Y. Heng Kee, A. Eriksson, and F. Neumann. Guaranteed outlier removal with mixed integer linear programs. In *Proc. 2016 Conf. Comput. Vision Pattern Recognition*, pages 5858–5866, 2016. 2

[8] O. Chum and J. Matas. Optimal randomized RANSAC. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(8):1472–1482, 2008. 2

[9] P. David, D. Dementhon, R. Duraiswami, and H. Samet. SoftPOSIT: simultaneous pose and correspondence determination. *Int. J. Comput. Vision*, 59(3):259–284, 2004. 2, 6, 7

[10] F. Dellaert, S. M. Seitz, C. E. Thorpe, and S. Thrun. Structure from motion without correspondence. In *Proc. 2000 Conf. Comput. Vision Pattern Recognition*, volume 2, pages 557–564. IEEE, 2000. 2

[11] O. Enqvist, E. Ask, F. Kahl, and K. Åström. Robust fitting for multiple view geometry. In *Proc. 2012 European Conf. Comput. Vision*, pages 738–751. Springer Berlin Heidelberg, 2012. 2

[12] O. Enqvist, E. Ask, F. Kahl, and K. Åström. Tractable algorithms for robust model estimation. *Int. J. Comput. Vision*, 112(1):115–129, 2015. 2

[13] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 1, 2, 7

[14] J. Fredriksson, V. Larsson, C. Olsson, and F. Kahl. Optimal relative pose with unknown correspondences. In *Proc. 2016 Conf. Comput. Vision Pattern Recognition*, pages 1728–1736. IEEE, 2016. 2

[15] W. E. L. Grimson. *Object Recognition by Computer: The Role of Geometric Constraints*. MIT Press, Cambridge, MA, USA, 1990. 2

[16] B. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nölle. Review and analysis of solutions of the three point perspective pose estimation problem. *Int. J. Comput. Vision*, 13(3):331–356, 1994. 1, 2

[17] R. I. Hartley and F. Kahl. Global optimization through rotation space search. *Int. J. Comput. Vision*, 82(1):64–79, 2009. 2, 4

[18] J. A. Hesch and S. I. Roumeliotis. A direct least-squares (DLS) method for PnP. In *Proc. 2011 Int. Conf. Comput. Vision*, pages 383–390. IEEE, 2011. 1, 2

[19] D. P. Huttenlocher and S. Ullman. Recognizing solid objects by alignment with an image. *Int. J. Comput. Vision*, 5(2):195–212, 1990. 1

[20] F. Jurie. Solution of the simultaneous pose and correspondence problem using Gaussian error model. *Computer Vision and Image Understanding*, 73(3):357–373, 1999. 3

[21] L. Kneip and P. Furgale. OpenGV: A unified and generalized approach to real-time calibrated geometric vision. In *Proc. 2014 Int. Conf. Robotics and Automation*, pages 1–8. IEEE, 2014. 6, 7, 8

[22] L. Kneip, H. Li, and Y. Seo. UPnP: An optimal O(n) solution to the absolute pose problem with universal applicability. In *Proc. 2014 European Conf. Comput. Vision*, pages 127–142. Springer, 2014. 1, 2

[23] L. Kneip, D. Scaramuzza, and R. Siegwart. A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. In *Proc. 2011 Conf. Comput. Vision Pattern Recognition*, pages 2969–2976. IEEE, 2011. 1, 2, 7

[24] L. Kneip, Z. Yi, and H. Li. SDICP: Semi-dense tracking based on iterative closest points. In *Proc. 2015 British Machine Vision Conference*, pages 100.1–100.12. BMVA Press, Sep. 2015. 1

[25] A. H. Land and A. G. Doig. An automatic method of solving discrete programming problems. *Econometrica: Journal of the Econometric Society*, pages 497–520, 1960. 2, 3

[26] V. Lepetit, F. Moreno-Noguer, and P. Fua. EPnP: An accurate O(n) solution to the PnP problem. *Int. J. Comput. Vision*, 81(2):155–166, 2009. 1, 2, 8

[27] H. Li. Consensus set maximization with guaranteed global optimality for robust geometry estimation. In *Proc. 2009 Int. Conf. Comput. Vision*, pages 1074–1080. IEEE, 2009. 2

[28] H. Li and R. Hartley. The 3D-3D registration problem revisited. In *Proc. 2007 Int. Conf. Comput. Vision*, pages 1–8. IEEE, 2007. 2, 3

[29] Y. Li, N. Snavely, D. Huttenlocher, and P. Fua. Worldwide pose estimation using 3D point clouds. In *Proc. 2012 European Conf. Comput. Vision*, pages 15–29. Springer-Verlag, 2012. 2

[30] Y. Li, N. Snavely, and D. P. Huttenlocher. Location recognition using prioritized feature matching. In *Proc. 2010 European Conf. Comput. Vision*, pages 791–804. Springer, 2010. 2

[31] W.-Y. Lin, L.-F. Cheong, P. Tan, G. Dong, and S. Liu. Simultaneous camera pose and correspondence estimation with motion coherence. *Int. J. Comput. Vision*, 96(2):145–161, 2012. 2

[32] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004. 1

[33] A. Makadia, C. Geyer, and K. Daniilidis. Correspondence-free structure from motion. *Int. J. Comput. Vision*, 75(3):311–327, 2007. 2

[34] E. Marchand, H. Uchiyama, and F. Spindler. Pose estimation for augmented reality: a hands-on survey. *IEEE Trans. Vis. Comput. Graphics*, 22(12):2633–2651, 2016. 1

[35] F. Moreno-Noguer, V. Lepetit, and P. Fua. Pose priors for simultaneously solving alignment and correspondence. In *Proc. 2008 European Conf. Comput. Vision*, pages 405–418. Springer, 2008. 2, 7

[36] J. L. Mundy. Object recognition in the geometric era: A retrospective. In J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors, *Toward Category-Level Object Recognition*, volume 4170 of *Lecture Notes in Computer Science*, pages 3–28. Springer, Berlin, Heidelberg, 2006. 1

[37] S. T. Namin, M. Najafi, M. Salzmann, and L. Petersson. A multi-modal graphical model for scene analysis. In *Proc. 2015 Winter Conf. Applications Comput. Vision*, pages 1006–1013. IEEE, 2015. 8

[38] T. Nöll, A. Pagani, and D. Stricker. Markerless Camera Pose Estimation - An Overview. In A. Middel, I. Scheler, and H. Hagen, editors, *Visualization of Large and Unstructured Data Sets - Applications in Geospatial Planning, Modeling and Engineering (IRTG 1131 Workshop)*, volume 19 of *OpenAccess Series in Informatics (OASIcs)*, pages 45–54, Dagstuhl, Germany, 2011. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. 1

[39] C. F. Olson. A general method for geometric feature matching and model extraction. *Int. J. Comput. Vision*, 45(1):39–54, 2001. 1

[40] C. Olsson, A. Eriksson, and R. Hartley. Outlier removal using duality. In *Proc. 2010 Conf. Comput. Vision Pattern Recognition*, pages 1450–1457. IEEE, 2010. 2

[41] C. Olsson, F. Kahl, and M. Oskarsson. Branch-and-bound methods for euclidean registration problems. *IEEE Trans. Pattern Anal. Machine Intelligence*, 31(5):783–794, 2009. 3

[42] D. P. Paudel, A. Habed, C. Demonceaux, and P. Vasseur. LMI-based 2D-3D registration: From uncalibrated images to Euclidean scene. In *Proc. 2015 Conf. Comput. Vision Pattern Recognition*. 2

[43] D. P. Paudel, A. Habed, C. Demonceaux, and P. Vasseur. Robust and optimal sum-of-squares-based point-to-plane registration of image sets and structured scenes. In *Proc. 2015 Int. Conf. Comput. Vision*, pages 2048–2056, 2015. 2

[44] T. Sattler, B. Leibe, and L. Kobbelt. Fast image-based localization using direct 2D-to-3D matching. In *Proc. 2011 Int. Conf. Comput. Vision*, pages 667–674. IEEE, 2011. 2

[45] T. Sattler, B. Leibe, and L. Kobbelt. Improving image-based localization by active correspondence search. In *Proc. 2012 European Conf. Comput. Vision*, pages 752–765. Springer-Verlag, 2012. 2

[46] K. Sim and R. Hartley. Removing outliers using the L_∞ norm. In *Proc. 2006 Conf. Comput. Vision Pattern Recognition*, volume 1, pages 485–494. IEEE, 2006. 2

[47] L. Svarm, O. Enqvist, F. Kahl, and M. Oskarsson. City-scale localization for cameras with known vertical direction. *IEEE Trans. Pattern Anal. Machine Intelligence*, 2016. 2

[48] L. Svärm, O. Enqvist, M. Oskarsson, and F. Kahl. Accurate localization and pose estimation for large 3D models. In *Proc. 2014 Conf. Comput. Vision Pattern Recognition*, pages 532–539. IEEE, 2014. 2

[49] J. Yang, H. Li, D. Campbell, and Y. Jia. Go-ICP: A globally optimal solution to 3D ICP point-set registration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(11):2241–2254, 2016. 2, 3, 4, 6

[50] J. Yu, A. Eriksson, T.-J. Chin, and D. Suter. An adversarial optimization approach to efficient outlier removal. In *Proc. 2011 Int. Conf. Comput. Vision*, pages 399–406. IEEE, 2011. 2