Parallax Bundle Adjustment on Manifold with Improved Global Initialization

Liyang Liu[1], Teng Zhang[1], Yi Liu[2], Brenton Leighton[1], Liang Zhao[1], Shoudong Huang[1], and Gamini Dissanayake[1]

[1] University of Technology Sydney, Sydney NSW 2007, Australia,
liyang.liu@uts.edu.au
[2] Huazhong University of Science and Technology, Wuhan 430074, China

Abstract. In this paper we present a novel extension to the parallax feature based bundle adjustment (BA). We take parallax BA into a manifold form (PMBA) along with an observation-ray based objective function. This formulation faithfully mimics the projective nature in a camera's image formation, resulting in a stable optimization configuration robust to low-parallax features. Hence it allows use of fast Dogleg optimization algorithm, instead of the usual Levenberg Marquardt. This is particularly useful in urban SLAM in which diverse outdoor environments and collinear motion modes are prevalent. Capitalizing on these properties, we propose a global initialization scheme in which PMBA is simplified into a pose-graph problem. We show that near-optimal solution can be achieved under low-noise conditions. With simulation and a series of challenging publicly available real datasets, we demonstrate PMBA's superior convergence performance in comparison to other BA methods. We also demonstrate, with the "Bundle Adjustment in the Large" datasets, that our global initialization process successfully bootstrap the full BA in mapping many sequential or out-of-order urban scenes.

Keywords: Bundle adjustment, Global SfM, Monocular SLAM

1 Introduction

Structure from Motion (SfM) as well as visual SLAM estimate 3D scene structures and camera poses simultaneously from 2D images. Bundle adjustment is the gold standard method in this activity that it finds the optimal pose and map in the least squares sense [1] to best explain the data. Solving such a non-linear least squares problem typically requires iterative Newton-based methods [2]: start with an initial guess, repetitively add increments by solving a normal equation until convergence. As shown in Table 1, this

Table 1. Three types of Newton-based methods

GN	LM	DL
$\triangle \mathbf{x} = \mathbf{H}^{-1}\mathbf{e}(\mathbf{x})$	$\triangle \mathbf{x} = (\mathbf{H} + \lambda \mathbf{I})^{-1}\mathbf{e}(\mathbf{x})$	$\triangle \mathbf{x} = (\lambda_1 \mathbf{H}^{-1} + \lambda_2 \mathbf{I})\mathbf{e}(\mathbf{x})$

approach comes in three forms: original Gauss-Newton (GN) when the equation is easy to solve (the Hessian matrix $\mathbf{H}$ has a small condition number), Levenberg Macquardt

(LM) as a damped GN if Hessian is near singular, and DogLeg (DL) as a combination of GN and the steepest descent method for fast convergence [7, 6]. Although all three methods can benefit from a sparse implementation, the overall convergence performances are different. Within each iteration step, LM tries to repetitively solve a new augmented equation (with an adjusted damping term λ) until error is reduced. In comparison, DL only inverts the Hessian once for all retries within an iteration, it is more efficient than LM in general. GN and DL are both considered risky for BA due to the large step size and are often avoided in practice. LM is a favourite of the robotics and computer vision communities for its safe handling despite its slowness [6].

Problematic Features In many modern BA systems [4, 11, 24], a 3D feature point is parameterized either as Euclidean coordinates (XYZ) or by the direction and inverse of depth from the first observing camera (IDP) [9]. A well-known problem for these representations is that existence of low parallax features during motion causes singularity in the Hessian matrix, a main contribution to GN divergence and numerical instability [9, 10]. A small change in error function leads to a large jump in the state variable, making it difficult to specify a consistent stop criterion. To avoid singularity, slow LM is commonly used for safe increment [7, 10] in place of GN or DL, efficiency is compromized for stability but could easily result in local minimum. These problematic features manifest in outdoor scenes as far away features and in street view scenes as features collinear to the observing camera motion. IDP can elegantly handle far features [9] but fail to cope with collinear ones [10]. See Fig. 1 and Fig. 8(a) for illustration of failure in conventional BA.

Robustness to problematic features is a major issue in urban SLAM. Several reme-

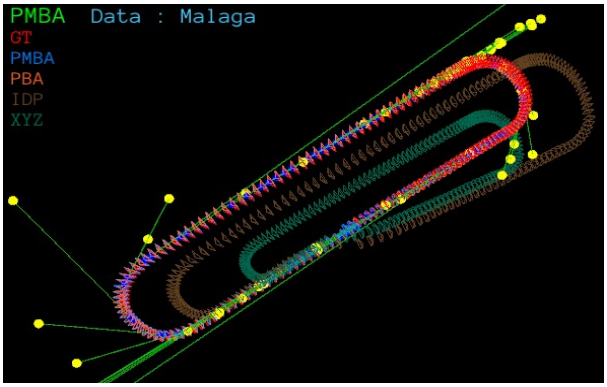


Fig. 1. Compare BA for "Malaga dataset": existence of collinear features (yellow dots) cause IDP (brown) and XYZ (green) to differ significantly from Ground Truth (red); PMBA (blue) and PBA (the original parallax-based BA [10], orange) do not encounter this issue. PMBA has fastest convergence, see Fig. 8(a).

dies are adopted to address it, with the common principle of separate treatment for problematic features and good ones. In ORB-SLAM [25], a prudent feature selection strategy is applied where features with in-sufficient parallax angles are discarded although they do contain some information. A hybrid method was proposed in [23], that

first estimates camera orientations with remote features then optimises with poses and near features. The vision smart factor proposed in [26] (implemented in GTSAM [24]) shares the same approach of [23]. It avoids degeneracy by using a flexible-sized error function. Recently [27] proposed a solution in which less weighting is given to the error terms for problematic features.

Our proposed solution in this paper takes a totally different viewpoints. After re-thinking the difference between state and state uncertainty, we argue that *the root cause for degeneracy is that the uncertainty of conventional feature forms is not uniformly bounded*. In our previous work [10], we presented the parallax angle based bundle adjustment (PBA) algorithm where a feature is represented by 3 highly observable angles without involving depth: a direction confining azimuth and elevation angles and a depth related parallax angle. [10] demonstrated that PBA is more robust and efficient compared to XYZ or IDP form BA's. Our proposed manifold formulation – PMBA is a continuation along this thinking and offers even better convergence properties.

Initialization Methods. BA due to its highly non-convex nature [2], requires good initial estimates to converge to global minimum. The common initialization methods are incremental or global. In incremental methods, with a simple start, many mid-level BAs are performed on each new pose insertion. This strategy draws the criticism that it is slow and relies heavily on picking good initial image pairs to progress. Example systems are VisualSFM [22], Bundler [12] and ORB-SLAM [25]. The alternative is the global strategy where all camera poses are initialised simultaneously. Global SfM thus bootstrapped shows higher efficiency and accuracy. The global strategy exposes many research challenges, and has been studied carefully in [17–19]. Our previous work [10] involved a simple initialization method that unfortunately is vulnerable to complicated camera motions and is only targeted at sequential inputs. The proposed initialization scheme in this paper addresses this issue.

Contributions and Paper Structure. This paper provides a novel BA formulation and initialization method robust to problematic features. First, we present (in Sect. 2) a novel BA formulation using parallax feature paramterization on manifold, its retraction method and an observation ray based error function. Next we show that the underlying optimization exhibits non-singular Hessians and bounded error functions, fully suppressing degeneracy due to problematic features. These good convergence properties allow fast DL optimization and is robust to urban scenes. In Sect. 3, we propose a global initialization strategy in which the PMBA problem is simplified into an easy-to-solve position registration problem. We show that the simplification leads to near-optimal solution under low-noise condition. We develop theorems and analysis for both contributions. In Sect. 4, we verify our claims through simulation and a series of large-scale publicly available datasets, all including low-parallax features. We present theorem proofs and reconstruction results in the supplementary material [3].

Notations. Throughout this paper, $\lfloor \mathbf{x} \rfloor_\times$ denotes a skew symmetric matrix from vector $\mathbf{x} \in \mathbb{R}^3$. $\xi()$ denotes the normalization operator, and $\xi(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ gives the direction of an arbitrary vector $\mathbf{x}$. We use the calligraphic font and roman font to represent manifold

and Euclidean variable respectively. For example, the i'th camera pose is represented as $\mathcal{T}_i = (\mathbf{R}_i, \mathbf{P}_i) \in \mathbb{SE}(3)$. For the j'th feature, $\mathbf{F}_j$ denotes its coordinates and $\mathcal{F}_j$ denotes its PMBA parameters. We use subscript $^{(l)}$ to indicate the reference frame is local.

2 Parallax Bundle Adjustment on Manifold

In this section, we introduce the PMBA formulation. We first define its feature parameterization in manifold domain, then show a retraction method and its compatible error function. Next we give a thorough theoretical analysis on the boundedness of its information matrix, hence proves its smooth convergence without singularity. All these factors lead to the possibility of using faster DL optimization method, which is a significant improvement over previous work [10].

2.1 Feature parameterization

A feature's depth information can be computed from the parallax between observations from different viewpoints.

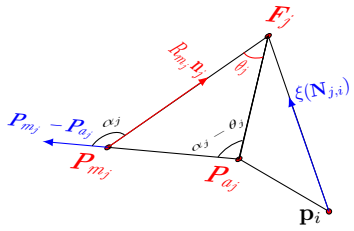


Fig. 2. Feature point $\mathbf{F}_j$ anchored by $\mathbf{P}_{m_j}$ and $\mathbf{P}_{a_j}$ with parallax angle θ_j. An arbitrary observing camera is shown at position $\mathbf{P}_i$. Directions of ray from $\mathbf{P}_{m_j}$ and $\mathbf{P}_i$ are labeled as $\mathbf{R}_{m_j}\mathbf{n}_j$ and $\xi(\mathbf{N}_{j,i})$ respectively, all in global frame.

For a 3D feature point $\mathbf{F}_j$, amongst the set of observing cameras $\mathbb{T}_j$, we choose a main anchor $\mathbf{T}_{m_j}$ and an associate anchor $\mathbf{T}_{a_j}$ that form the best parallax angle from their observation rays. This geometric relationship for feature j is illustrated in Fig. 2.

In manifold domain, we denote the feature as $\mathcal{F}_j$ and over-parameterize it by its unit observation ray vector $\mathbf{n}_j$ in main-anchor frame, and the parallax angle θ_j,

$$\mathcal{F}_j = (\cos\theta_j, \sin\theta_j, \mathbf{n}_j) \tag{1}$$

This parameterization only defines the relative structure of the feature with respect to its two anchors. The scale of point $\mathbf{F}_j$ is implicitly defined by the relative translation of the two anchors, and is computed as

$$\mathbf{F}_j = \overrightarrow{P_{a_j}P_{m_j}} + \mathbf{P}_{m_j} = \frac{\sin(\alpha_j - \theta_j)}{\sin(\theta_j)} \|\mathbf{P}_{m_j} - \mathbf{P}_{a_j}\| \mathbf{R}_{m_j}\mathbf{n}_j + \mathbf{P}_{m_j} \tag{2}$$

where

- $\mathbf{n}_j \in \mathbb{R}^3$ is the direction of observation ray from $\mathbf{P}_{m_j}$ to $\mathbf{F}_j$, local in $\mathcal{T}_{m_j}$ frame.
- $\frac{\sin(\alpha_j - \theta_j)}{\sin(\theta_j)} \|\mathbf{P}_{m_j} - \mathbf{P}_{a_j}\|$ is the distance between the two, from sine rule.
- $\mathbf{R}_{m_j}$ is the rotation for main anchor frame $\mathcal{T}_{m_j}$.
- α_j is the angle between vector $(\mathbf{P}_{m_j} - \mathbf{P}_{a_j})$ and $\mathbf{R}_{m_j}\mathbf{n}_j$, see (11) for derivation

Remark 1. In the original PBA parameterization [10], ray direction $\mathbf{n}_j$ was defined by an elevation and azimuth angle in the global frame, camera's orientation $\{\mathbf{R}_i\}$ in Euler angles. When a feature's elevation angle is $\frac{\pi}{2}$, its azimuth angle becomes irrelevant. Expressing directions in sinusoids of angles is a potential source of singularity. In PMBA, both $\mathbf{n}_j$ and $\mathbf{R}_i$ are in the manifold domain. Moreoever, $\mathbf{n}_j$ is newly defined to be in $\mathcal{T}_{m_j}$'s local frame, for ease of multi-camera system application.

2.2 State variable retraction in manifold

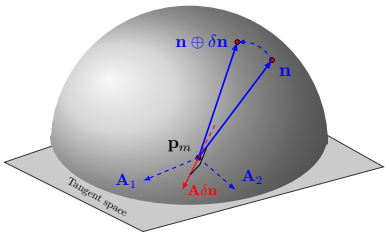


Fig. 3. Retraction of ray $\mathbf{n}$ in main anchor

Optimization in manifold follows a three step procedure [8]: lift a manifold variable to its tangent space, solve a normal equation to obtain the Euclidean increment, and retract back to manifold. We use the pose retraction method described in [20]. For feature ray direction, we give a natural definition of uncertainty as a normally distributed rotational perturbation to the directional vector as shown in Fig. 3. The rotation's axis constitutes a plane normal to the ray passing through the observing camera, the plane is the tangent space. We can express the perturbed ray direction as:

$$\tilde{\mathbf{n}}_j = \mathrm{Exp}(\mathbf{A}_{\mathbf{n}_j}\delta\mathbf{n}_j)\mathbf{n}_j, \qquad \delta\mathbf{n}_j \in \mathcal{N}(0,\Sigma). \tag{3}$$

where $\delta\mathbf{n}_j \in \mathbb{R}^2$, $\mathbf{A}_{\mathbf{n}_j} \in \mathbb{R}^{3\times 2}$ is the left null-space of $\mathbf{n}_j$ and $[\mathbf{A}_{\mathbf{n}_j}\ \mathbf{n}_j] \in \mathbb{SO}(3)$, $\mathrm{Exp}()$ is the exponential map for $\mathbb{SO}(3)$. The optimal perturbation is the increment for retraction:

$$\mathcal{F}_j \boxplus \delta\mathcal{F}_j = (\cos(\theta_j + \delta\theta_j), \sin(\theta_j + \delta\theta_j), \mathrm{Exp}(\mathbf{A}_{\mathbf{n}_j}\delta\mathbf{n}_j)\mathbf{n}_j). \tag{4}$$

where the total increment $\delta\mathcal{F}_j = \left[\delta\theta_j, \delta\mathbf{n}_j\right] \in \mathbb{R}^3$ has same dimensionality as conventional parameterization.

2.3 Error function and optimization formulation

In PMBA, we estimate camera poses $\boldsymbol{\mathcal{T}} = \{(\mathbf{R}_i, \mathbf{P}_i)\}_{i=1,\cdots,M}$ and feature parameters $\boldsymbol{\mathcal{F}} = \{\mathcal{F}_j \in \mathbb{M}^3\}_{j=1,\cdots,N}$ from a set of images $\{I_i\}$. When the feature j is observed from the pose $\mathcal{T}_i$, the monocular sensor intercepts the light ray $\mathbf{N}_{j,i}$ that passes through its centre to the feature point at the image pixel $\mathbf{u}_{m_{j,i}}$.

In conventional bundle adjustment, pixel imprints are used directly as measurements to estimate feature set $\mathbb{F} = \{\mathbf{F}_j \in \mathbb{R}^3\}_{j=1,\cdots,N}$ together with poses. As a maximum a posterior (MAP) problem, the conventional BA is formulated as:

$$\min_{\boldsymbol{\mathcal{T}},\mathbb{F}} \sum_{i\in\mathbb{T}_j,j} \|e_{ij}(\boldsymbol{\mathcal{T}}_i, \mathbf{F}_j)\|^2 = \min_{\boldsymbol{\mathcal{T}},\mathbb{F}} \sum_{i\in\mathbb{T}_j,j} \|\mathbf{K}\circ\pi(\mathbf{R}_i^{\intercal}(\mathbf{F}_j - \mathbf{P}_i)) - \mathbf{u}_{m_{j,i}}\|^2 \tag{5}$$

Where $\mathbf{K}$ represents the camera calibration matrix and π is the homogeneous normalization operator. The error function $e_{ij}(\cdot) \in \mathbb{R}^2$ does not give clue to cheirality property (lies in front of or behind a given camera).

On the other hand, the observation ray $\mathbf{N}_{j,i} \in \mathbb{R}^3$ includes directional information and should provide a better measurement for feature source than its 2D pixel counterpart. From (2), we express $\mathbf{N}_{j,i}$ (in global frame) as:

$$\begin{aligned} \mathbf{N}_{j,i} &= \sin(\theta_j)(\mathbf{F}_j - \mathbf{P}_i) \\ &= \sin(\alpha_j - \theta_j)\|\mathbf{P}_{m_j} - \mathbf{P}_{a_j}\|\mathbf{R}_{m_j}\mathbf{n}_j + \sin(\theta_j)(\mathbf{P}_{m_j} - \mathbf{P}_i) \end{aligned} \tag{6}$$

Note that we have applied a factor of $\sin(\theta_j)$ for convenience of mathematical manipulation. The variable forms the observation ray appears in this paper is in Table 2.

Table 2. Various forms of observation ray in this paper

Global ray	Global ray direction	Local ray	Local ray direction
$\mathbf{N}_{j,i} = \mathbf{F}_j - \mathbf{P}_i$	$\xi(\mathbf{N}_{j,i}) = \frac{\mathbf{F}_j - \mathbf{P}_i}{\|\mathbf{F}_j - \mathbf{F}_i\|}$	$\mathbf{N}^{(l)}_{j,i} = \mathbf{R}_i^\intercal(\mathbf{F}_j - \mathbf{P}_i)$	$\xi(\mathbf{N}^{(l)}_{j,i}) = \frac{\mathbf{R}_i^\intercal(\mathbf{F}_j - \mathbf{P}_i)}{\|\mathbf{R}_i^\intercal(\mathbf{F}_j - \mathbf{P}_i)\|}$

We now define the ray direction based error function, essentially a "chordal distance" of bearing vectors (on the sphere):

$$\mathbf{e}_{ij}(\mathbf{N}^{(l)}_{j,i}) := \mathbf{v}_{j,i} - \xi(\mathbf{N}^{(l)}_{j,i}) \in \mathbb{R}^3, \tag{7}$$

where $\mathbf{v}_{j,i} = \xi(\mathbf{K}^{-1}\mathbf{u}_{m_{j,i}}) \in \mathbb{R}^3$ is the observation ray's measured direction.

Moving this measurement to the global frame, expressing all states in manifold, we come to the final non-linear least squares problem for PMBA:

$$\min_{\boldsymbol{\mathcal{X}}} \|f(\boldsymbol{\mathcal{X}})\|^2 = \min_{\boldsymbol{\mathcal{T}},\boldsymbol{\mathcal{F}}} \sum_{i\in\mathbb{T}_j, j} \|\xi(\mathbf{N}^{(l)}_{j,i}) - \mathbf{R}_i\mathbf{v}_{j,i}\|^2, \qquad \boldsymbol{\mathcal{X}} = (\boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{F}}) \tag{8}$$

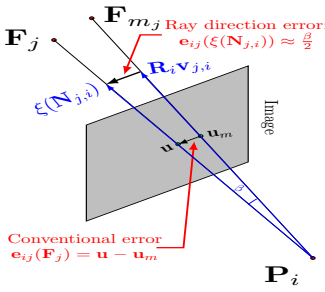


Fig. 4. Camera measurement error

The difference between pixel and ray-direction error forms are shown in Fig. 4.

Remark 2. The conventional 2D error function (5) shows ambiguity for frontal or hinder feature positions, the non-uniqueness of function values brings many local minimums and saddle points. In comparison, the proposed 3D error function (7) leads to improved monotonicity with reduced local minima. Its magnitude can be modelled with the angle β between the estimated and measured ray direction : $\|\mathbf{e}_{ij}\| = 2\sin(\frac{\beta}{2}) \in [0..2]$. This covers the entire range of small error ($\beta \to 0$) to behind-scene error ($\beta \to \pi$). Further, (5) employs homogeneous normalization $\pi(\cdot)$ which prevents feature's local Z-ordinate from approaching zero, causing BA discontinuity. (7) employs vector normalization $\xi(\cdot)$ instead, is almost totally continuous.

With the above discussion, the new optimization (8) should exhibit significantly increased convergence region and the ability to correct feature state from many erroneous estimates through successive iterations, including behind the camera case.

2.4 Theoretical analysis on convergence properties

We now give an analysis on PMBA's convergence properties.
Consider the Hessian matrix of the problem (8)

$$\mathbf{H} = \mathbf{J}^\intercal \mathbf{J} = \begin{bmatrix} \mathbf{H_{TT}} & \mathbf{H_{TF}} \\ \mathbf{H_{TF}^\intercal} & \mathbf{H_{FF}} \end{bmatrix}, \tag{9}$$

where $\mathbf{J} := \frac{\partial f(\boldsymbol{\mathcal{X}} \boxplus \Delta\boldsymbol{\mathcal{X}})}{\partial \Delta\boldsymbol{\mathcal{X}}}|_{\Delta\boldsymbol{\mathcal{X}}=\mathbf{0}}$ and $\mathcal{X} \boxplus \Delta\boldsymbol{\mathcal{X}} := (\boldsymbol{\mathcal{T}} \boxplus \Delta\boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{F}} \boxplus \Delta\boldsymbol{\mathcal{F}})$. Like the Hessian matrix in conventional BA, $\mathbf{H_{FF}}$ is block diagonal. Applying the *Schur's omplement* method, the dominant computation in each Newton method's iteration boils down to solving the following normal equation:

$$(\mathbf{H_{TT}} - \mathbf{H_{TF}}\mathbf{H_{FF}^{-1}}\mathbf{H_{TF}^\intercal})\Delta\boldsymbol{\mathcal{T}} = -\left[\mathbf{I}\ \mathbf{H_{TF}}\mathbf{H_{FF}^{-1}}\right] f(\boldsymbol{\mathcal{X}}), \tag{10}$$

In conventional BA, existence of problematic features causes the matrix $\mathbf{H_{TT}} - \mathbf{H_{TF}}\mathbf{H_{FF}^{-1}}\mathbf{H_{TF}^\intercal}$ and the block matrix $\mathbf{H_{FF}}$ (with slight abuse of notation) to be ill-conditioned at the neighborhood of global minimum. The global minimum locates at a "long flat valley" [10] such that solvers fail or require large number of iterations to converge, see Fig. 8*(a)* for illustration.

In comparison, PMBA's formulation (8), thanks to the re-defined retraction (4) and the compatible error function (7), faithfully complies with projective geometry in image formation, *is therefore well-posed with significantly improved local observability despite of "problematic" features*.

Theorem 1. *Under the formulation (8),* $\mathbf{H_{FF}}$ *is consistently non-singular for any* $\boldsymbol{\mathcal{X}}$ *and* $\mathbf{H_{FF}} \geq \mathbf{I}$.

Proof. See [3] for proof.

Theorem 1 completely suppresses all ill-conditioned $\mathbf{H_{FF}}$ so normal equations (GN) are always solvable. With increased convergence region, the DL trust-region method can be safely used to increase iteration step size (a sum of GN step plus steepest descent), and minimize iteration time (cheap retries). Theorem 1 can also be appreciated from an Information Theory perspective: the Hessian matrix at global minimum is the inverse of the covariance matrix (up to a scale) and thus the uncertainty of the parallax angle θ_j and the direction $\mathbf{n}_j$ is uniformly bounded.

Remark 3. The original PBA [10] cannot guarantee non-singularity in $\mathbf{H_{FF}}$ due to use of standard addition retraction for feature, Euler angles for orientation and the error function (5).

Remark 4. Although the matrices $\mathbf{H_{TT}}$ and $\mathbf{H_{TF}}$ are denser, compared to those in XYZ or IDP, $\mathbf{H_{TT}} - \mathbf{H_{TF}}\mathbf{H_{FF}^{-1}}\mathbf{H_{TF}^\intercal}$ shows same sparsity. Thus the computational time for each iteration in PMBA is comparable to conventional BA, see [10].

3 Global Initialization

In this section, we derive a initialization strategy compatible to PMBA. The goal is to derive camera poses from a set of essential matrices. We do this in three steps. We first identify well-matched image pairs to obtain their epipolar geometry's (EG). From the EG-pairs we then initialize rotations and features. Finally we simplify the original PMBA problem into a pose-graph problem and estimate camera positions with two optimization stages: constrained least square (CLS) and non-linear (NLS) optimisation. We show that near-optimal solution can be obtained under low-noise conditions. This pipeline of global initialization and PMBA are illustrated in Fig. 5.

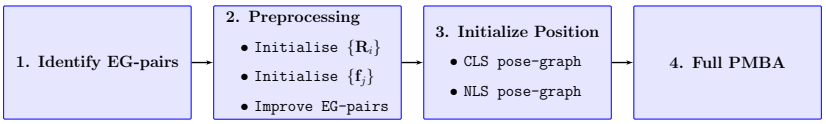


Fig. 5. Full Global Initialization + PMBA pipeline.

3.1 Orientation and feature initialization

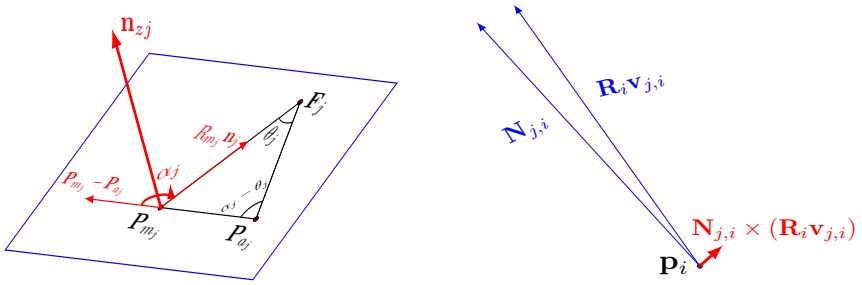
Following the approach in [16–18], we first obtain an initial guess for orientations. This requires a maximal set of EG-pairs (relative rotation and translation) be formed from two-view matches. We use Kneip's 5-point algorithm [14] to calculate each EG's essential matrix. Next we build a maximum spanning tree from EG-pairs and discard bad pairs thus establishing accurate image connectivity. We choose the tree root as our reference frame, and use tree branches to help form prior rotation estimates. This is especially useful with out-of-order image inputs. We use the state-of-art chordal initialization [13] for rotation averaging. We found the output rotations very reliable and can feed them back to the tree for outlier-pruning and relative translation fine-tuning [15].

With accurate rotation estimates, we are ready for feature initialization. We adopt the default anchor selection strategy in [10], with the small change that co-visible poses that participate in anchor selection have to be part of an EG-pair. This step ensures best as-can-be parallax angle be given to each feature point. We stress that any problematic features corresponding to low parallax angles do stay in the state and do not affect convergence under PMBA. Good features together with problem ones work together to shape the final solution.

Remark 5. In PMBA, feature paramterization does not require scale information. Their initialization therefore relies only on camera rotations [10] and are very accurate. We thus completely avoid unreliable/expensive linear triangulation.

3.2 Position initialization

Rotations and features initialized above are highly accurate, they can be assumed fixed thus do not participate in the subsequent optimization. This way PMBA is transformed



(*a*) Linearize ray function: rotate $(\mathbf{P}_{a_j} - \mathbf{P}_{m_j})$ by $(\pi - \alpha_j)$ about $\mathbf{n}_{zj}$ becomes $\|\mathbf{P}_{m_j} - \mathbf{P}_{a_j}\|\mathbf{R}_{m_j}\mathbf{n}_j$.

(*b*) PMBA reformatted as a constrained -LS problem: minimize cross product

Fig. 6. Simplification of PMBA into position only problem

into a pose-graph problem where the unknown variables are positions only. This pose-graph problem it not convex but can be further simplified. We do so first by approximating the ray $\mathbf{N}_{j,i}$ from a non-linear function of poses to a linear combination of positions. Specifically, the non-linear term $\|\mathbf{P}_{m_j} - \mathbf{P}_{a_j}\|\mathbf{R}_{m_j}\mathbf{n}_j$ in (6) can be seen as a rotation of $(\mathbf{P}_{a_j} - \mathbf{P}_{m_j})$ to $\mathbf{R}_{m_j}n_j$ about axis $\mathbf{n}_{zj}$ by angle $\pi - \alpha_j$, as illustrated in Fig. 6(*a*). Both $\mathbf{n}_{zj}$ and α_j are locally observable and are computed as,

$$\alpha_j = \arccos(\xi(\mathbf{P}_{m_j} - \mathbf{P}_{a_j})^\intercal(\mathbf{R}_{m_j}\mathbf{n}_j)), \qquad \mathbf{n}_{zj} = \xi(\mathbf{P}_{a_j} - \mathbf{P}_{m_j}) \times (\mathbf{R}_{m_j}\mathbf{n}_j) \quad (11)$$

We now give the linearized expression for the observation ray, denoted $\bar{\mathbf{N}}_{j,i}$:

$$\begin{aligned}\bar{\mathbf{N}}_{j,i}(\mathbf{P}_{m_j}, \mathbf{P}_{a_j}, \mathbf{P}_i) = \\ \sin(\bar{\alpha}_j - \bar{\theta}_j)\exp(\bar{\mathbf{n}}_{zj}(\pi - \bar{\alpha}_j))(\mathbf{P}_{a_j} - \mathbf{P}_{m_j}) + \sin(\bar{\theta}_j)(\mathbf{P}_{m_j} - \mathbf{P}_i)\end{aligned} \quad (12)$$

After substituting $\bar{\mathbf{N}}_{j,i}$ into (8), we establish a "*position only*" optimization,

$$\min_{\mathbf{P}} h(\mathbf{P}, \bar{\mathbf{R}}, \bar{\mathbf{F}}) := \min_{\{\mathbf{P}_i\}} \sum_{i \in \mathbb{T}_j, j} \|\xi(\bar{\mathbf{N}}_{j,i}(\mathbf{P}_{m_j}, \mathbf{P}_{a_j}, \mathbf{P}_i)) - \bar{\mathbf{R}}_i\mathbf{v}_{j,i}\|^2. \quad (13)$$

This approach of position registration from unitized direction vectors is inspired by the non-linear method from [18]. The cost-function in [18] is essentially an algebraic difference of inter-pose directions. Whereas ours is based on pose-feature directions without solving for the features, and can directly handle collinear motions by virtue of parallax structure (see Sect. 3.3).

Remark 6. Considering (13) is still a nonlinear problem, an initial guess is needed by its iterative solver. We obtain the initial values by computing the optimum of a constrained least squares problem. The objective is to minimize the cross-product between ray $N_{j,i}$ (linear to positions) and $\mathbf{R}_i\mathbf{v}_{j,i}$ as shown in Fig. 6*(b)*. Due to sign ambiguity in cross-products, we add the linear constraint to ensure cheirality condition. The overall CLS problem is:

$$\min_{\{\mathbf{P}_i\}} \sum_{i \in \mathbb{T}_j, j} \|\lfloor \bar{\mathbf{R}}_i\mathbf{v}_{j,i} \rfloor_\times \bar{\mathbf{N}}_{j,i}(\mathbf{P}_{m_j}, \mathbf{P}_{a_j}, \mathbf{P}_i)\|^2, \quad z(\bar{R}_i\bar{\mathbf{N}}_{j,i}) >= 0. \quad (14)$$

3.3 Theoretical analysis

Theorem 2. *With accurate initial estimate for orientation, the formulation (13) can always converge to a near-optimal solution for the BA problem (8).*

Proof. See [3] for proof.

Theorem 2 proves the correctness and robustness of the proposed initialization. Moreover, from the viewpoint of computational complexity the pose-graph problem (13) exhibits much reduced variable size than (8) and the expensive feature retraction operation is also avoided.

It is interesting that Theorem 2 provides a theoretical assurance that partitioning the BA problem into a pose-graph initialization step and a full BA step is a sound approach. In fact, the *separation* strategy is well known in SLAM systems. In [29] a SLAM problem with range and bearing observations is shown to exhibit a separable structure: given orientation, robot and feature positions are linear in the corresponding error function. The separable structure is further exploited in [30] to form an efficient iterative solver with better convergence. Undoubtedly, visual SLAM is far more complex where depth information is not readily observable. Our proposed initialization as well as other algorithms [17–19] all intrinsically apply the *separation* strategy to simplify the complex BA problem.

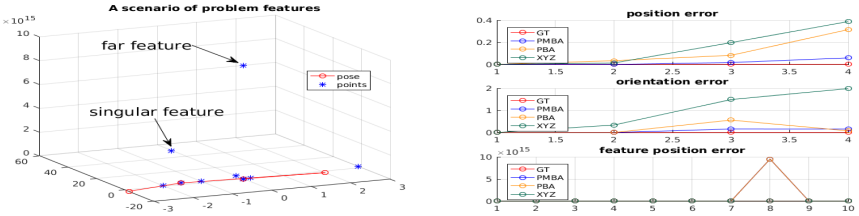
Remark 7. Here we do not claim the proposed global initialization is the best one but it is very compatible to PMBA. The pose-graph problem includes all feature observations in its objective function, hence contain sufficient information to handle collinear motions. Further, it does not require strong triplet image association as in [17].

Remark 8. Note that the proposed method is friendly to robust methods such as pseudo Huber, L^1-norm or outlier detection technique. Further, the non-linear model is still formulated in a probabilistic framework, different from the "Linear Global Translation Estimation" reported in [19].

4 Evaluation on PMBA performance

4.1 Simulation

We demonstrate PMBA's ability to handle problematic features with a simple simulation test. The scene consists of 4 poses and 10 features, two of which are problematic, as shown in Fig. 7*(a)*. One problematic feature is a far feature, the other initialized with values that would cause singularity in the original PBA algorithm. The BAs under comparison are: XYZ-BA, PBA and PMBA. We run 4 iterations for each BA and collect their Hessian's. At the end we gather BA estimates deviation from ground truth. The results are listed in Table 3. One can see that PMBA has good Hessian condition numbers and accurate optimized estimates, PBA and XYZ-BA show consistently large condition numbers and high errors. This confirms Theorem 1 that This can be explained by our Theorem 1 that the Hessian in PMBA does not exhibit Hessian singularity yet other BAs can.



(*a*) Simulation with problem features

(*b*) Comparison of the error from PMBA, PBA and XYZ-BA

Fig. 7. Compare three BA forms in a simulated scene with problem features

Table 3. Comparison of $\mathbf{H_{FF}}$'s condition number during optimization and final state error for PMBA, PBA and XYZ-BA

Convergence Properties	PMBA	PBA	XYZ-BA
Iter-0 $cond(\mathbf{H_{FF}})$	9.74	1.46E+4	1.22E+94
Iter-1 $cond(\mathbf{H_{FF}})$	5.68	1.46E+4	1.22E+94
Iter-2 $cond(\mathbf{H_{FF}})$	8.80	1.46E+4	1.22E+94
Iter-3 $cond(\mathbf{H_{FF}})$	5.74	1.46E+4	3.53E+95
Final χ^2_{error}	2.58E-3	5.37E-2	3.43E-2

4.2 Large dataset test

We conducted a series of real datasets to compare performance of the proposed PMBA (8) and original PBA, IDP and XYZ, aiming to address following questions:

- **Robustness.** With degeneracy scenario disappearing, can DL be safely used in PMBA implementation?
- **Efficiency.** If DL were applied for PMBA, how fast can the optimization be?
- **Accuracy.** Since the PMBA formulation employs a different error function (7). Is the global minimum accurate?

All methods are tested against six very challenging datasets, which are also accessible from OpenSLAM[3]. In particular,

- *Fake-pile* is collected by the Google tango tablet in normal lab environment with a fake bridge pile in the middle, showing close and far features.
- *Malaga* [21] is a classic street view dataset. It is collected using an electric car equipped camera facing the road, consisting of many collinear features.
- *Village* and *College* are aerial photogrammetric datasets. The low feature to observation ratio implies existence of many small parallax features
- *Usyd-Mainquad-2* and *Victoria-cottage* are collected at University of Sydney campus, full of far features. See [3] for reconstruction results.

[3] https://svn.openslam.org/data/svn/ParallaxBA/

Table 4. Comparison of convergence performance for PMBA, PBA, XYZ-BA, IDP-BA

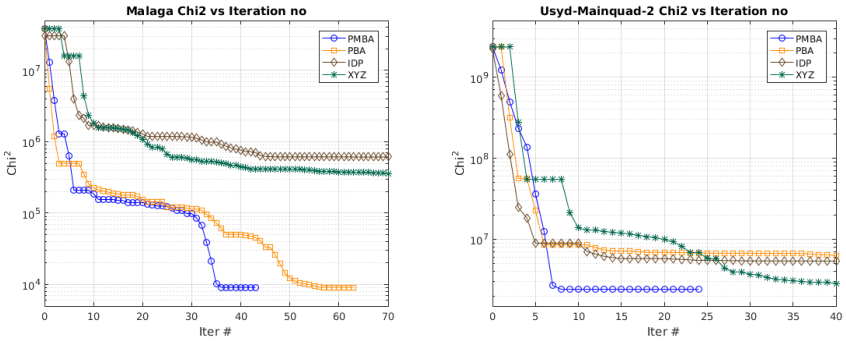
Dataset	Type	# Pose / # Feat / # Obsv	# Equation solve / # Iteration	Final Chi2	Time $[sec]$
Fake-pile	PMBA	135	**9 / 9**	**1.7E+2**	**0.7**
	PBA	/12,741	23 / 23	1.7E+2	1.9
	IDP	/53,878	104 / 102	1.7E+2	6.0
	XYZ		116 / 108	1.2E+3	4.7
Malaga	PMBA	170	**44 / 31**	**9.1E+3**	**21.6**
	PBA	/305,719	64 / 47	9.1E+3	35.4
	IDP	/779,268	230 / 170	5.8E+5	93.8
	XYZ		110 / 85	3.3E+5	39.0
Village	PMBA	90	**12 / 12**	3.3E+4	**31.8**
	PBA	/305,719	13 / 13	**3.3E+4**	36.0
	IDP	/779,268	19 / 19	3.3E+4	35.2
	XYZ		18 / 18	3.3E+4	26.3
College	PMBA	468	33 / 33	1.1E+6	**334.4**
	PBA	/1,236,502	**31 / 31**	**1.1E+6**	370.5
	IDP	/3,107,524	34 / 34	1.1E+6	255.3
	XYZ		295 / 193	1.0E+7	1361.0
Victoria cottage	PMBA	400	**19 / 16**	1.1E+6	**70.5**
	PBA	/153,632	88 / 66	1.2E+6	301.4
	IDP	/890,057	49 / 48	**1.1E+6**	157.9
	XYZ		47 / 44	1.2E+6	124.3
Usyd -Mainquad	PMBA	424	**25 / 25**	**2.4E+6**	**214.5**
	PBA	/227,615	101 / 57	3.6E+6	642.6
	IDP	/1,607,082	301 / 191	4.6E+6	1994.7
	XYZ		76 / 58	2.8E+6	423.7

We set all BAs from the same starting point use the imperfect initialization method from [10] to observe iteration behaviour. We find that PBA, IDP and XYZ show unstable behaviour under DL. PMBA, in comparison, has always worked well with DL. This can be explained by our prediction that PMBA has a large convergence region and is consistently well-posed. We therefore list DL results for PMBA and LM for all other BA's.

We implement all BAs in C++ and use Ceres-solver [4] as the optimization engine. All BAs are tested on an Intel-i7 CPU running one thread. We use ray direction cost function for PMBA, and compute its corresponding uv-based Chi2 error at each iteration step with current estimate, to compare with other BAs on a common error metric. This scheme is not fair for PMBA, yet is the only convincing way to evaluate performance amongst all methods. Despite of this treatment, we found PMBA the best performer in all tests, consistent with our expectation.

Selected convergence plots are shown in Fig. 8, more can be found in [3]. All collected metrics are summarized in Table 4.

Further, for the Malaga dataset which is full of problematic features (Fig. 1), we observe that the PMBA estimates and Ground Truth are very close, yet conventional



(*a*) Malaga (*b*) Usyd-Mainquad-2

Fig. 8. Convergence plots for PMBA, PBA, IDP and XYZ

BA gives significant error. This is also seen in Table 4, conventional BAs converge to a local minimum, whereas both PBA and PMBA can converge to their respective global minimums. Figure 8 and Table 4 confirm that the error function (7) is practical, consistent with the claim in [28]. In conclusion, these experiments all give positive answers to the raised questions.

4.3 Evaluation of PMBA global initialization

Finally, we conduct tests to verify our initialization strategy. We use datasets from the "Bundle Adjustment in the Large" (BAL) datebase[4] [5] and the datasets in Sect. 4.2. We implement a PMBA-based SfM pipeline complying to the procedure in Fig. 5 in C++, using Ceres [4] as the optimization engine. For comparison, we run same tests on an incremental pipeline similar to Bundler, also written in C++ using Ceres [4].

These datasets are selected for showing street scene (Ladybug-1370), diverse proximity scene (Trafalgar-126, Venice-427) or photometric aerial scene (College), all exposing challenges for conventional BA. Since camera calibration is beyond the scope of this activity, we apply the reported optimal camera settings from BAL and PBA websites and only test undistorted versions of these data. We stress that our initial pose and feature values are purely generated from the rotation averaging and translation registration methods described in Sect. 3, without using the initial values provided by [5].

The performance comparison results are shown in Table 5. Here the column labeled "Ours" is the proposed CLS-NLS-PMBA pipeline, the column labeled "Incremental" refers to the incremental pipleline. Both incremental and our pipeline give similar outputs, we therefore only tabulate the timing information. Figure 9 illustrates our pipeline results in blue, red color shows BAL results or PBA results from [10] for the College dataset. The red and blue data are almost identical. We also give detailed reconstruction results at various stages of our pipeline in [3].

[4] http://grail.cs.washington.edu/projects/bal/

Table 5. Complete pipleline comparison

			number of BA's		time[min]	
Dataset	order	num poses	Ours	Incremental	Ours	Incremental
Ladybug-1370	sequential	1370	1	394	3.33	65
Trafalgar-126	out-of-order	126	1	25	1.01	1.1
Venice-427	out-of-order	427	1	49	6.62	17.4
College	sequential	468	1	238	9.63	85.43

The results in Table 5 shows that our global SfM pipeline uses less computation time and BA invocations than the incremental method in all tests. This result together with the pipeline output plots in [3] confirm our proposed initialization strategy is viable.

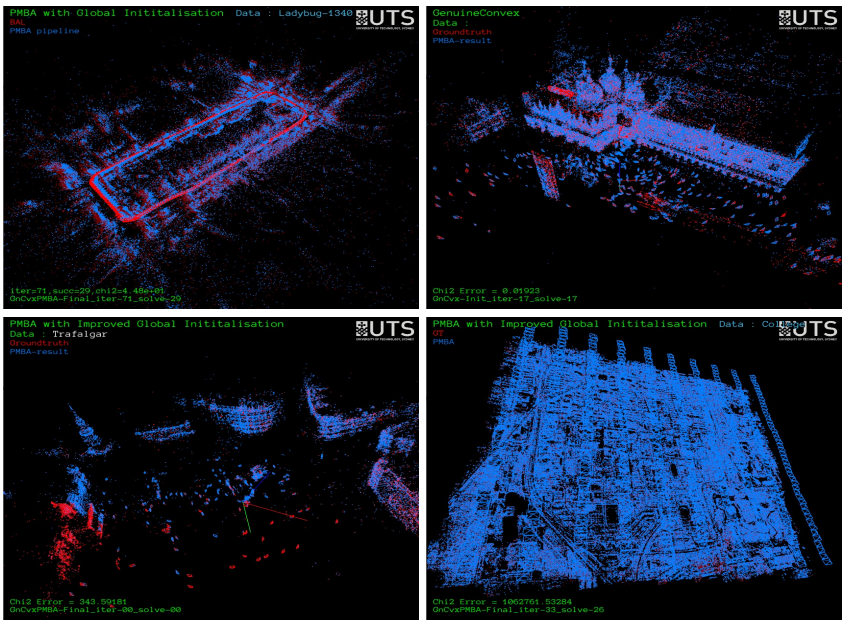


Fig. 9. Reconstruction results of full PMBA pipeline on test datasets

5 Conclusion

In this work, we proposed a new bundle adjustment formulation – PMBA which utilizes parallax angle based on-manifold feature parametrization and observation-ray based objective function. We proved that under the new formulation the ill-conditioned cases due to problematic features can be avoided without any manual intervention, which results in much better convergence and robustness properties.

Furthermore, motivated by the separable structure in the visual SLAM problem and ease of parallax feature initialization, we derived a novel global initialization process for

PMBA. We use a simplified pose-graph model that can guarantee a near-optimal solution to bootstrap the original BA problem. Experimental results show that the proposed initialization can provide efficient and accurate estimates and is a viable global initialization strategy for many challenging situations including sequential and out-of-order images.

The promising results of the global initialization plus PMBA pipeline using publicly available datasets demonstrate that the proposed technique can deal with different challenging data. In the future, we are planning to integrate the proposed pipeline with efficient visual SLAM front-end to develop a robust and efficient SfM system.

References

1. Triggs, B., McLauchlan, P., Hartley, R., Fitzgibbon, A.: Bundle Adjustment – A Modern Synthesis. In: Triggs, W., Zisserman, A., Szeliski, R. (eds) Vision Algorithms: Theory and Practices, vol. 1883, pp. 298-375. Springer, Heidelberg (2000). doi: 10.1007/3-540-44480-7_21
2. Cadena., C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.: Past, present, and future of simultaneous localization and mapping: Towards the robust-perception age. IEEE Transactions on Robotics, 32(6), 1309-1332. IEEE Press (2016). doi: 10.1109/TRO.2016.2624754
3. L. Liu, T. Zhang, Y. Liu, B. Leighton, L. Zhao, Huang, S., Dissanayake, G.: Supplementary Material to: Parallax bundle adjustment on manifold with improved global initialization. University of Technology Sydney, 2018.
4. Agarwal, S., Mierle, K.: Ceres Solver. http://ceres-solver.org
5. Agarwal, S., Snavely, N., Seitz, S. M., Szeliski R.: Bundle Adjustment in the large. In: Daniilidis K., Maragos P., Paragios N. (eds) Computer Vision ECCV 2010, vol. 6312, pp. 29-42. Springer, Berlin, Heidelberg (2010). http://grail.cs.washington.edu/projects/bal
6. Rosen, D.M., Kaess, M., Leonard, J.J.: RISE: An incremental trust-region method for Robust online sparse least-squares estimation. IEEE Transactions on Robotics, 30(5), 1091-1108. IEEE Press (2014). doi: 10.1109/TRO.2014.2321852
7. Lourakis, M. L. and Argyros, A. A.: Is Levenberg-Marquardt the most efficient optimization algorithm for implementing bundle adjustment? In: Tenth IEEE International Conference on Computer Vision (ICCV'05). IEEE Press, Beijing (2005). doi: 10.1109/ICCV.2005.128
8. Smith, S. T.: Trust-region methods on Riemannian manifolds. Foundations of Computational Mathematics, vol. 7, no. 3, pp. 303-330. Springer (2007). doi: 10.1007/s10208-005-0179-9
9. Civera, J., Davison, A., Martinez Montiel, J.: Inverse depth parametrization for monocular slam. IEEE Transactions on Robotics, 24(5), 932-945. IEEE Press (2008). doi: 10.1109/TRO.2008.2003276
10. Zhao, L., Huang, S., Sun, Y., Yan, L., Dissanayake, G.: ParallaxBA: bundle adjustment using parallax angle feature parametrization. The International Journal of Robotics Research, 34(4-5), 493-516. SAGE Publications (2015). doi: 10.1177/0278364914551583
11. Kuemmerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: G2o: A general framework for graph optimization. In: 2011 IEEE International Conference on Robotics and Automation (ICRA). IEEE Press (2011). doi: 10.1109/ICRA.2011.5979949
12. Snavely, N., Seitz, S., Szeliski, R.: Photo Tourism: Exploring image collections in 3D. ACM Transactions on Graphics (TOG), vol. 25, no. 3, pp. 835-846. ACM, New York (2006). doi: 10.1145/1141911.1141964
13. Carlone, L., Tron, R., Daniilidis, K., Dellaert, F.: Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization. In: 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE Press, Seattle (2015). doi: 10.1109/ICRA.2015.7139836

14. Kneip, L., Siegwart, R., Pollefeys, M.: Finding the Exact Rotation between Two Images Independently of the Translation. In: Fitzgibbon A., Lazebnik S., Perona P., Sato Y., Schmid C. (eds) Computer Vision ECCV 2012, vol. 7577, pp. 696-709. Springer, Berlin, Heidelberg (2012). doi: 10.1007/978-3-642-33783-3_50
15. Kneip, L., Furgale, P.: OpenGV: A unified and generalized approach to real-time calibrated geometric vision. In: 2014 IEEE International Conference on Robotics and Automation (ICRA). IEEE Press, Hong Kong (2014). doi: 10.1109/ICRA.2014.6906582
16. Crandall, D., Owens, A., Snavely, N., Huttenlocher, D.: Discrete-continuous optimization for large-scale structure from motion. In: International Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Press, Colorado (2011). doi: 10.1109/CVPR.2011.5995626
17. Jiang, N., Cui, Z., Tan, P.: A Global Linear Method for Camera Pose Registration. In: 2013 IEEE International Conference on Computer Vision (ICCV). IEEE Press, Sydney (2013). doi: 10.1109/ICCV.2013.66
18. Wilson, K., Snavely, N.: Robust Global Translations with 1DSfM. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision ECCV 2014, vol. 8691, pp. 61-57. Springer, Cham (2014). doi: 10.1007/978-3-319-10578-9_5
19. Cui, Z., Jiang, N., Ping, T.: Linear Global Translation Estimation from Feature Tracks. In: Proceedings of the British Machine Vision Conference (BMVC), pp. 46.1-46.13. BMVA Press, (2015). doi: 10.5244/C.29.46
20. Zhang, T., Wu, K., Song, J., Huang, S., Dissanayake, G.: Convergence and Consistency Analysis for a 3-D Invariant-EKF SLAM. IEEE Robotics and Automation Letters, 2(2), 733-740. IEEE Press (2017). doi: 10.1109/LRA.2017.2651376
21. Blanco, J., Moreno, F., Gonzalez, J.: A collection of outdoor robotic datasets with centimeter-accuracy ground truth. J. Auton Robot, vol. 27, no. 4, pp. 327. Springer, US (2009). doi: 10.1007/s10514-009-9138-7
22. Wu, C.: VisualSFM: A visual structure from motion system. http://ccwu.me/vsfm/
23. Zhang, H., Hasith, K., Wang, H.: A hybrid feature parametrization for improving stereo-SLAM consistency. In: 2017 13th IEEE International Conference on Control Automation (ICCA). IEEE Press, Ohrid (2017). doi: 10.1109/ICCA.2017.8003201
24. Dellaert, F.: Factor graphs and GTSAM: A hands-on introduction. Georgia Institute of Technology. https://bitbucket.org/gtborg/gtsam/
25. Mur-Artal, R., Montiel, M., Tardós, J.: ORB-SLAM: a Versatile and Accurate Monocular SLAM System. IEEE Transactions on Robotics, 31(5), 1147–1163. IEEE Press, (2015). doi: 10.1109/TRO.2015.2463671
26. Carlone, L., Kira, Z., Beall, C.: Eliminating conditionally independent sets in factor graphs: A unifying perspective based on smart factors. In: 2014 IEEE International Conference on Robotics and Automation (ICRA). IEEE Press, Hong Kong (2014). doi: 10.1109/ICRA.2014.6907483
27. Lee, C., Yoon, K.: Exploiting Feature Confidence for Forward Motion Estimation. CoRR, vol. abs/1704.07145 (2017). http://arxiv.org/abs/1704.07145
28. Im, S., Ha, H., Rameau, F., Jeon, H., Choe, G., Kweon, I.: All-Around Depth from Small Motion with a Spherical Panoramic Camera. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision ECCV 2016, vol. 9907, pp. 156–172 (2016). Springer, Cham (2016). doi: 10.1007/978-3-319-46487-9_10
29. Huang, S. Lai, Y., Frese, U., Dissanayake, G.: How far is SLAM from a linear least squares problem? In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE Press, Taipei (2010). doi: 10.1109/IROS.2010.5652603
30. Khosoussi, K., Huang, S., Dissanayake, G.: A Sparse Separable SLAM Back-End. IEEE Transactions on Robotics, 32(6), 1536-1549. IEEE Press (2016). doi: 10.1109/TRO.2016.2609394