Parallax Bundle Adjustment on Manifold with Convexified 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. The new model is convex in nature and can guarantee a near-optimal solution. 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) / visual SLAM estimates 3D scene structures and camera poses simultaneously from 2D images. Bundle adjustment is the gold standard method of SfM, in that it finds 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 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]. LM is a favourite of the robotics and computer vision communities for its safe handling despite its slowness [6]. GN and DL are both considered risky for BA due to the large step size and are often avoided.

Table 1. Three types of Newton-based methods

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

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 compromised 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 handle far features elegantly [9] but fail to cope with collinear ones [10]. See Fig. 1 and Fig. 8(a) for illustration of failure in conventional BA.

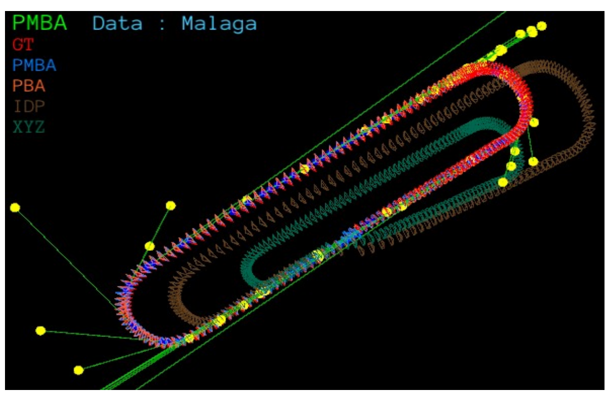


Fig. 1. Compare BA for “Malaga dataset”: existence of collinear features (yellow dots) cause IDP (brown) and XYZ (green) to differ significantly from Ground Truth (red); PMBA (blue) and PBA (the original parallax-based BA [10], orange) do not encounter this issue. PMBA has fastest convergence, see Fig.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

Robustness to problematic features is a major issue in urban SLAM. Several remedies 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 degenerate cases by using a flexible-size error function. Recently [27] proposed a solution in which less weighting is given to the error terms for problematic features.

In our previous work [10], we presented the parallax based bundle adjustment (PBA) algorithm that treats the problem with a totally different viewpoint. [10] showed that the root reason for degeneracy is that conventional feature paramterisations causes BA's covariance to be unbounded. A PBA feature is represented by three angles (elevation, azimuth and parallax) and can define a 3D point without involving depth. [10] demonstrated that PBA is more robust and efficient compared to BA's in XYZ or IDP form. 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 include incremental or global. In incremental strategy, with a simple start, many mid-level BAs are performed on each new pose insertion. The incremental 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 strategy is global initialization where all camera poses are initialised simultaneously. Global SfM thus bootstrapped shows higher efficiency and accuracy. The strategy exposes many research challenges, and has been studied carefully in [17][18][19]. Our previous work [10] suggested a simple initialization method that unfortunately is vulnerable to complicated camera motions and is only targeted at sequential inputs. The proposed PMBA pipeline 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 Section 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 allows fast DL optimization and is robust to urban scenes. In Section 3, we propose a global initialization strategy in which the PMBA problem is simplified into an easy-to-solve position registration problem. The initializaiton process consists of a constrained least squares step and a convex optimization step. We show that the convex formulation can guarantee a near-optimal solution. We develop theorems and analysis for both contributions. In Section 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, $S(\mathbf{x})$ denotes a skew symmetric matrix from vector $\mathbf{x} \in \mathbb{R}^3$, and $S(\mathbf{x})\mathbf{y} = \mathbf{x} \times \mathbf{y}$. We use the term $\mathbf{T}_i = (\mathbf{R}_i, \mathbf{p}_i) \in \mathbb{SE}(3)$ to represent the camera pose at time-step i. Subscript $^{(l)}$ indicates frame is local. Decoration $\breve{}$ indicates normalized vector, for example, $\breve{\mathbf{N}}_{j,i} = \frac{\mathbf{N}_{j,i}}{\|\mathbf{N}_{j,i}\|}$.

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. We also show the error function is bounded and globally continuous. All these factors lead to the possibility of using faster DL optimization method. This 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.

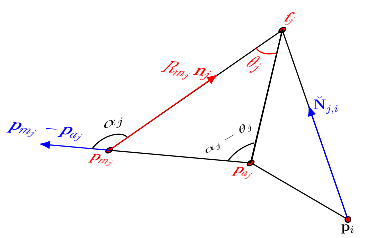


Fig. 2. Feature $\mathbf{f}_j$ anchored by $\mathbf{p}_{m_j}$ and $\mathbf{p}_{a_j}$ with parallax angle θ_j. An arbitrary observing camera is shown at position $\mathbf{p}_i$. Directions of ray from $\mathbf{p}_{m_j}$ and $\mathbf{p}_{i_j}$ are labeled as $\mathbf{R}_{m_j}\mathbf{n}_j$ and $\breve{N}_{j,i}$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

For a feature $\mathbf{f}_j$, 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, the feature $\mathbf{f}_j$ can be over-parameterized by its unit observation ray vector $\mathbf{n}_j$ in main-anchor frame, and the parallax angle θ_j,

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

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

$$\mathbf{f}_j = d_j \mathbf{R}_{m_j}\mathbf{n}_j + \mathbf{p}_{m_j} = \frac{\sin(\alpha_j - \theta_j)}{\sin(\theta_j)} \|\mathbf{p}_{m_j} - \mathbf{p}_{a_j}\| \mathbf{R}_{m_j}\mathbf{n}_j + \mathbf{p}_{m_j} \tag{2}$$

where

- $d_j = \frac{\sin(\alpha_j - \theta_j)}{\sin(\theta_j)} \|\mathbf{p}_{m_j} - \mathbf{p}_{a_j}\|$ is the local depth of the feature j in the main anchor frame, from sine rule.
- $\mathbf{R}_{m_j}$ is the rotation for main anchor frame $\mathbf{T}_{m_j}$.
- $\mathbf{n}_j \in \mathbb{R}^3$ is the direction of observation ray from $\mathbf{p}_{m_j}$ to $\mathbf{f}_j$, local in $\mathbf{T}_{m_j}$.
- α_j is the angle between vector $(\mathbf{p}_{m_j} - \mathbf{p}_{a_j})$ and $\mathbf{R}_{m_j}\mathbf{n}_j$, see (12) 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. Expressing direction 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 $\mathbf{T}_{m_j}$'s local frame, for ease of multi-camera system application.

2.2 State variable retraction in manifold

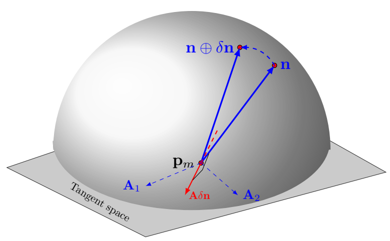


Fig. 3. Retraction of ray $\mathbf{n}$ in main anchor

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 pose retraction 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, and is the tangent space, summarized in the following equation:

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

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

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

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

2.3 Error function and optimization formulation

The visual SLAM problem estimates camera poses $\mathbb{T} = \{(\mathbf{R}_i, \mathbf{p}_i)\}_{i=1,\cdots,M}$ and feature positions $\mho = \{\mathbf{f}_j \in \mathbb{R}^3\}_{j=1,\cdots,N}$ from a set of images $\{I_i\}$. We denote the alternative on-manifold feature set as $\mathbb{F} = \{\mathbf{F}_j \in \mathbb{M}^3\}_{j=1,\cdots,N}$.

When the feature j is observed from the pose $\mathbf{T}_i$, the monocular sensor intercepts the light ray $\mathbf{N}_{j,i}$ that passes through its centre to the feature point at the image pixel $\mathbf{u}_{m_{j,i}}$. Using the measurement information, the maximum a posterior (MAP) problem for poses and points is formed:

$$\min_{\mathbb{T},\mho} \sum_{i\in\mathbb{T}_j, j} \|e_{ij}(\mathbf{T}_i, \mathbf{f}_j)\|^2 \tag{5}$$

Conventionally, the error function $e_{ij}(\cdot)$ is given by:

$$e_{ij}(\mathbf{f}_j) := \mathbf{K} \circ \pi(\mathbf{R}_i^\intercal(\mathbf{f}_j - \mathbf{p}_i)) - \mathbf{u}_{m_{j,i}} \quad \in \mathbb{R}^2. \tag{6}$$

Where $\mathbf{K}$ represents camera calibration matrix and π is the homogenous normalization operator. The observation ray $\mathbf{N}_{j,i} \in \mathbb{R}^3$ includes direction information and should be a better measurement of feature source. From (2), we express $\mathbf{N}_{j,i}$ in global frame as:

$$\mathbf{N}_{j,i} = \sin(\theta_j)(\mathbf{f}_j - \mathbf{p}_j) = \sin(\alpha_j - \theta_j)\|\mathbf{p}_{m_j} - \mathbf{p}_{a_j}\|\mathbf{R}_{m_j}\mathbf{n}_j + \sin(\theta_j)(\mathbf{p}_{m_j} - \mathbf{p}_i). \quad (7)$$

Note that we have applied a factor of $\sin(\theta_j)$, for convenience of mathematical manipulation. With abuse of notation, we use the same symbol $\mathbf{N}_{j,i}$. Thoughout this paper, the observation ray comes in many different forms, as listed in Table 2.

Table 2. Various forms of observation ray in this paper

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

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

$$\mathbf{e}_{ij}(\breve{\mathbf{N}}^{(l)}_{j,i}) := \breve{\mathbf{v}}_{j,i} - \breve{\mathbf{N}}^{(l)}_{j,i} \in \mathbb{R}^3, \quad (8)$$

where $\breve{\mathbf{v}}_{j,i} = \frac{\mathbf{K}^{-1}\mathbf{u}_{m_{j,i}}}{\|\mathbf{K}^{-1}\mathbf{u}_{m_{j,i}}\|} \in \mathbb{R}^3$ is the measured directional vector for the feature j in the pose $\mathbf{T}_i$.

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

$$\min_{\mathcal{X}} \|f(\mathcal{X})\|^2 = \min_{\mathbb{T},\mathbb{F}} \sum_{i \in \mathbb{T}_j, j} \|\breve{\mathbf{N}}_{j,i} - \mathbf{R}_i\breve{\mathbf{v}}_{j,i}\|^2, \qquad \mathcal{X} = (\mathbb{T}, \mathbb{F}) \quad (9)$$

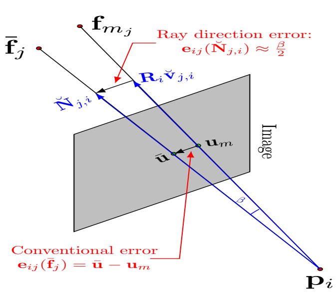


Fig. 4. Camera measurement error functions

The difference between conventional and ray error forms are shown in Fig. 4.

Remark 2. The conventional error function (6) does not enforce cheirality requirement (visible points should lie in front of the cameras), thus causing many local minima and saddle points. The error function (8) is globally continuous and its derivative is bounded, unlike (6). Its dimensionality is extended to 3D from 2D, meaning direction information also contributes to the Hessian.

We notice that the error equation (8) implies the residual $\|\mathbf{e}_{ij}\| = 2\sin(\frac{\beta}{2})$, where β is the angle between the estimated and measured ray direction. Thus, the error equation is bounded. In contrast to conventional 2D cost functions, our error function (8) operates in 3D and thus can handle the case when the feature point lies behind the observing camera.

2.4 Theoretical analysis on convergence properties

We now give an analysis on PMBA's convergence properties.
Consider the Hessian matrix of the problem (9)

$$\mathbf{H} = \mathbf{J}^\intercal \mathbf{J} = \begin{bmatrix} \mathbf{H_{TT}} & \mathbf{H_{TF}} \\ \mathbf{H}^\intercal_{\mathbf{TF}} & \mathbf{H_{FF}} \end{bmatrix}, \tag{10}$$

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

$$(\mathbf{H_{TT}} - \mathbf{H_{TF}}\mathbf{H}^{-1}_{\mathbf{FF}}\mathbf{H}^\intercal_{\mathbf{TF}})\Delta\mathbf{T} = -\begin{bmatrix} \mathbf{I} & \mathbf{H_{TF}}\mathbf{H}^{-1}_{\mathbf{FF}} \end{bmatrix} f(\mathcal{X}), \tag{11}$$

In conventional BA, existence of problematic features causes the matrix $\mathbf{H_{TT}} - \mathbf{H_{TF}}\mathbf{H}^{-1}_{\mathbf{FF}}\mathbf{H}^\intercal_{\mathbf{TF}}$ 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 (9), thanks to the re-defined retraction (4) and the error function (8), faithfully complies with projective geometry in computer vision, results in an well-behaved Hessian, can therefore fully avoid the ill-conditioned cases caused by "problematic" features.

Theorem 1. *Under the formulation (9),* $\mathbf{H_{FF}}$ *is consistently non-singular for any* $\mathcal{X}$ *and* $\mathbf{H_{FF}} \geq \mathbf{I}$.

Proof. See [3] for proof.

Theorem 1 completely suppresses all ill-conditioned $\mathbf{H_{FF}}$ so convergence becomes easily achievable. As a result, DL can be safely used to improve efficiency. One can also appreciate Theorem 1 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 (6).

Remark 4. Although the matrices $\mathbf{H_{TT}}$ and $\mathbf{H_{TF}}$ are denser, compared to those in XYZ or IDP, $\mathbf{H_{TT}} - \mathbf{H_{TF}}\mathbf{H}^{-1}_{\mathbf{FF}}\mathbf{H}^\intercal_{\mathbf{TF}}$ 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 novel initialization strategy. The goal is to produce a unique camera pose solution given a set of essential matrices. We do this in three steps. We first identify matchable epipolar geometry (EG) pairs. Using this results, we then perform rotation and feature initialization. To estimate camera positions, we instead solve for a simplified pose-graph problem, which is done in a quadratic programming stage and convex optimization stage. We prove that a near-optimal solution can be obtained by this strategy. This pipeline of global initialization and final BA are illustrated in Fig. 5.

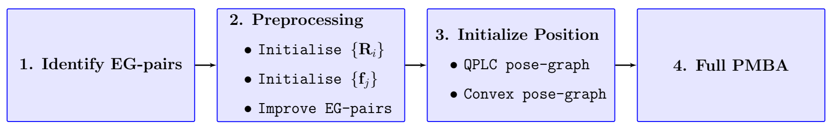


Fig. 5. Full Global Initialization + PMBA pipeline.

3.1 Orientation and feature initialization

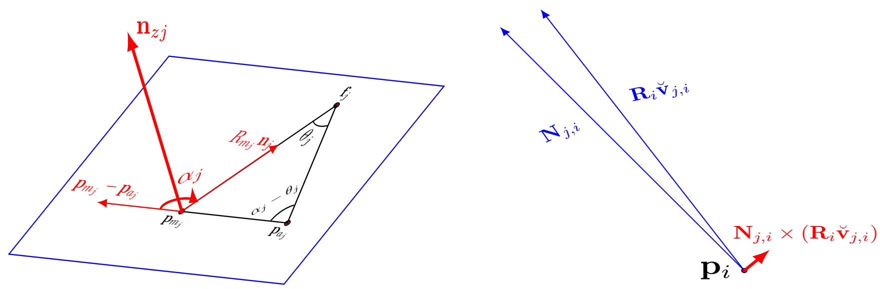
Following the approach in [16][17][18], we first compute an initial guess for orientation. This requires a maximal set of EG-pairs (relative rotation and translation) be computed from two-view matches. We apply Kneip's 5-point algorithm [14] for our input calculation. Next we build a maximum spanning tree from EG-pairs to build image connectivity and select good pairs. The connected network helps to build prior values before rotation averaging starts. This is especially useful in out-of-order image inputs. The tree root is chosen as the reference frame. We use the state-of-art chordal initialization [13] for rotation estimation. We found the output rotations very reliable and can feed them back to the tree for outlier-pruning and relative translation fine-tuning [15].

Having obtained accurate estimates for orientations and EG-pairs, we are ready to perform feature initialization. The default anchor selection strategy was given in [10]. We use the same algorithm for anchor selection, 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 given thus do not participate in the subsequent optimization. Thus PMBA is transformed into a pose-graph problem where the unknown variables are positions only. This pose-graph problem is non-convex but can be further improved. We do so first by approximating



(*a*) Linearize observation ray: $\|\mathbf{p}_{m_j} - \mathbf{p}_{a_j}\|\mathbf{R}_{m_j}\mathbf{n}_j$ is $(\mathbf{p}_a - \mathbf{p}_m)$ rotated about $\mathbf{n}_{zj}$ by $(\pi - \alpha_j)$.

(*b*) PMBA reformatted as a QPLC problem: minimize cross product

Fig. 6. Simplification of PMBA into pose-graph model

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 (7) 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,

$$\alpha_j = \arccos\Big(\frac{(\mathbf{p}_{m_j} - \mathbf{p}_{a_j})^\intercal(\mathbf{R}_{m_j}\mathbf{n}_j)}{\|\mathbf{p}_{m_j} - \mathbf{p}_{a_j}\|}\Big), \qquad \mathbf{n}_{zj} = \frac{(\mathbf{p}_{a_j} - \mathbf{p}_{m_j})}{\|\mathbf{p}_{a_j} - \mathbf{p}_{m_j}\|} \times (\mathbf{R}_{m_j}\mathbf{n}_j) \quad (12)$$

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

$$\bar{\mathbf{N}}_{j,i} = \sin(\bar{\alpha}_j - \bar{\theta}_j)\exp(\bar{\mathbf{n}}_{zj}(\pi - \bar{\alpha}_j))(\mathbf{p}_a - \mathbf{p}_m) + \sin(\bar{\theta}_j)(\mathbf{p}_{m_j} - \mathbf{p}_i), \quad (13)$$

After substituting $\bar{\mathbf{N}}_{j,i}$ into (9), we establish a "*position only*" optimization,

$$\min_{\{\mathbf{p}\}} h(\mathbf{p}, \bar{\mathbb{R}}, \bar{\mathbb{F}}) := \min_{\{\mathbf{p}\}} \sum_{i \in \mathbb{T}_j, j} \|\breve{\bar{\mathbf{N}}}_{j,i} - \bar{\mathbf{R}}_i\check{\mathbf{v}}_{j,i}\|^2. \quad (14)$$

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 our optimization is based on pose-feature directions without computing the features, by virtue of parallax structure it can directly handle collinear motions and is convex – see Section 3.3 for detailed analysis.

Remark 6. Considering (14) is still a nonlinear problem, an initial guess is needed to solve (14). Since $\bar{\mathbf{N}}_{j,i}$ is linear to positions, we further simplify (14) to a linearly constrained Quadratic Programming (QPLC) problem: to minimize the cross-product between ray $N_{j,i}$ and $\mathbf{R}_i\check{\mathbf{v}}_{j,i}$ as shown in Fig. 6*(b)*. Due to sign ambiguity in cross-products, we add the linear constraint to ensure cheirality condition, i.e.,

$$\min_{\{\mathbf{p}\}} \sum_{i \in \mathbb{T}_j, j} \|S(\bar{\mathbf{R}}_i\check{\mathbf{v}}_{j,i})\bar{\mathbf{N}}_{j,i}\|^2, \quad z(\bar{R}_i\bar{\mathbf{N}}_{j,i}) >= 0. \quad (15)$$

3.3 Theoretical analysis

Theorem 2. *With accurate initial estimate for orientation, the formulation (14) can always converge to a near-optimal solution for the BA problem (9). Furthermore, the problem (14) is convex when EG pairs are noise-free.*

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 (14) exhibits much reduced variable size than (9) 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, separation strategy is a long held view in SLAM. In [29] a SLAM problem with range and bearing observations is shown to exhibit a separable structure: given orientation, camera and feature positions are linear in the corresponding error function. The *separation* strategy is also discussed in great length in [30] with a modified iterative solver. Undoubtedly, BA is far more complex where depth information is not readily observable. Our global initialization as well as the initialization algorithms in [17][18][19] all intrinsically apply the *separation* strategy to tackle the BA problem.

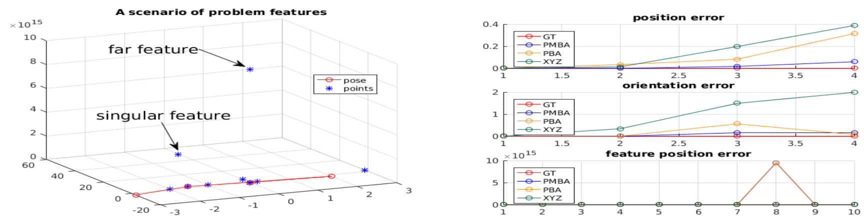
Remark 7. Here we do not claim the proposed global initialization is the best one but it is very compatible to PMBA. The pose-graph problem includes all feature observations in its objective function, hence contain sufficient information to handle collinear motions. Further, it does not require strong 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, this convexified 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. We run 4 iterations for BAs under comparison and collect their Hessian's. At the end we gather their estimates deviation to ground truth. The results are listed in Table 3. We can see that PMBA has normal condition numbers and good optimized estimates, PBA and XYZ-BA show consistently large condition numbers and high errors. This confirms our prediction that PMBA has well-behaved Hessians during optimization.



(*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 (9) and original PBA, IDP and XYZ, aiming to address following questions:

- **Robustness.** With degeneracy scenario disappearing, can DL be safely used in PMBA?
- **Efficiency.** If DL were safely applied for PMBA formulation, how fast can the optimization process be?
- **Accuracy.** Since the PMBA formulation employs a different error function (8). 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

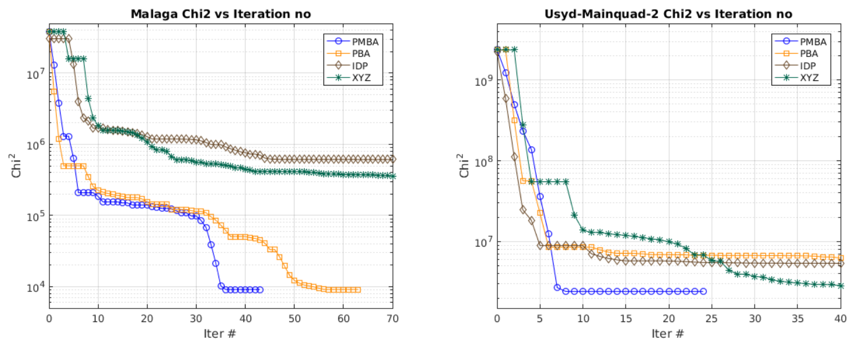
Dataset	Test-type	# Pose / # Feat / # Obsv	# Equation solving / # Iteration	Final Chi2	Time[sec]
Fake-pile	PMBA	135	**9 / 9**	**1.7E+2**	**0.7**
	PBA	/12,741	23 / 23	1.7E+2	1.9
	IDP	/53,878	104 / 102	1.7E+2	6.0
	XYZ		116 / 108	1.2E+3	4.7
Malaga	PMBA	170	**44 / 31**	**9.1E+3**	**21.6**
	PBA	/305,719	64 / 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 Theorem 1 that the Hessian in PMBA does not exhibit singularity yet other BAs can. 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 BA gives significant error. This is also seen in Table 4, conventional BAs converge



(*a*) Malaga

(*b*) Usyd-Mainquad-2

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

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 (8) is practical, consistent with the claim in [28]. In conclusion, these experiments all give positive answers to the raised questions.

4.3 Evaluation of convexified 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 Section 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 Section 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 QPLC-Convex-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.

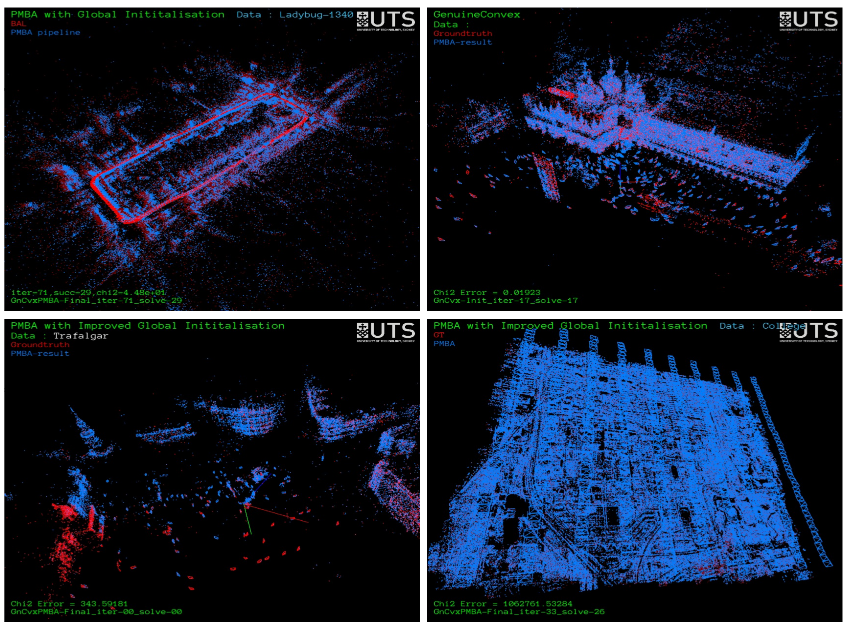


Fig. 9. Reconstruction results of full PMBA pipeline on test datasets

5 Conclusion

In this work, we proposed a new bundle adjustment formulation (PMBA) which utilizes parallax angle based feature parametrization on manifold and observation-ray based objective function. We proved that under the new formulation the ill-conditioned cases due to problematic features can be theoretically 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, 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) (2016) 1309-1332. 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 convexified initialization. Submitted to WAFR 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: European Conference on Computer Vision (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) (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) Volume 1, pp. 1526-1531. doi: 10.1109/ICCV.2005.128
8. Smith, S. T.: Trust-region methods on Riemannian manifolds. Foundations of Computational Mathematics 7(3) (2007). pp. 303-330
9. Civera, J., Davison, A., Martinez Montiel, J.: Inverse depth parametrization for monocular slam. IEEE Transactions on Robotics 24(5) (2008). pp. 932-945. 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) (2015) 493-516. doi: 10.1177/0278364914551583
11. Kuemmerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: G2o: A general framework for graph optimization. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on, IEEE (2011), pp.3607-3613. doi: 10.1109/ICRA.2011.5979949
12. Snavely, N., Seitz, S., Szeliski, R.: Photo Tourism: Exploring image collections in 3D. Proceedings of ACM Transactions on Graphics (SIGGRAPH 2006).
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). pp. 4597-4604. 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: Proceedings of the 12th European Conference on Computer Vision (ECCV) 2012. (2012). pp. 696-709. 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), 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) 2011. pp. 3001-3008. 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) 2013. pp. 481-488. doi: 10.1109/ICCV.2013.66
18. Wilson, K., Snavely, N.: Robust Global Translations with 1DSfM. In: Proceedings of the European Conference on Computer Vision (ECCV) 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) 2015. pp. 46.1-46.13. 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 2017 2(2) 733-740. doi: 10.1109/LRA.2017.2651376
21. Blanco, J., Moreno, F., Gonzalez, J.: A collection of outdoor robotic datasets with centimeter-accuracy ground truth. Autonomous Robots 2009, pp. 327. 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: Proceedings of the 13th IEEE International Conference on Control Automation (ICCA) 2017. pp. 1021-1026. 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) (2015) 1147–1163. 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) 2014. pp. 4290-4297. 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: Proceedings of 14th European Conference of Computer Vision (ECCV) 2016, Part III. pp. 156–172. 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: Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, (2010) 3011-3016. 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) (2016). pp. 1536-1549. doi: 10.1109/TRO.2016.2609394