Revisiting Rolling Shutter Bundle Adjustment: Toward Accurate and Fast Solution

Bangyan Liao[1,2,*] Delin Qu[1,3,*] Yifei Xue[4] Huiqing Zhang[1] Yizhen Lao[1,†]
[1]College of Computer Science and Electronic Engineering, Hunan University
[2]College of Electrical and Information Engineering, Hunan University
[3]Shanghai AI Laboratory
[4]Jiangxi Provincial Natural Resources Cause Development Center

Abstract

We propose an accurate and fast bundle adjustment (BA) solution that estimates the 6-DoF pose with an independent RS model of the camera and the geometry of the environment based on measurements from a rolling shutter (RS) camera. This tackles the challenges in the existing works, namely, relying on high frame rate video as input, restrictive assumptions on camera motion and poor efficiency. To this end, we first verify the positive influence of the image point normalization to RSBA. Then we present a novel visual residual covariance model to standardize the reprojection error during RSBA, which consequently improves the overall accuracy. Besides, we demonstrate the combination of ***N****ormalization and covariance standardization* ***W****eighting in* ***RSBA*** *(NW-RSBA) can avoid common planar degeneracy without the need to constrain the filming manner. Finally, we propose an acceleration strategy for NW-RSBA based on the sparsity of its Jacobian matrix and Schur complement. The extensive synthetic and real data experiments verify the effectiveness and efficiency of the proposed solution over the state-of-the-art works.*

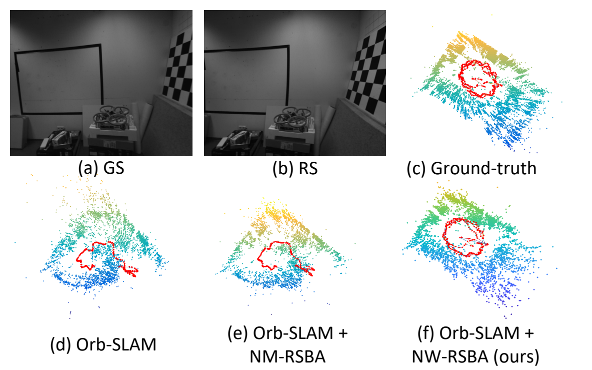

Figure 1. Images were captured at the same time with fast motion with (a) global shutter and (b) rolling shutter in sequence 10 of the TUM-RSVI dataset [21]. (c) classical Orb-SLAM [18] with GS input. (d) classical Orb-SLAM [18] with RS input. (e) Orb-SLAM with *NM-RSBA* [2]. (f) Orb-SLAM with proposed *NW-RSBA*.

1. Introduction

Bundle adjustment (BA) is the problem of simultaneously refining the cameras' relative pose and the observed points' coordinate in the scene by minimizing the reprojection errors over images and points [8]. It has made great success in the past two decades as a vital step for two 3D computer vision applications: structure-from-motion (SfM) and simultaneous localization and mapping (SLAM).

The CMOS camera has been widely equipped with the rolling shutter (RS) due to its inexpensive cost, lower energy consumption, and higher frame rate. Compared with the CCD camera and its global shutter (GS) counterpart, RS camera is exposed in a scanline-by-scanline fashion. Consequently, as shown in Fig. 1(a)(b), images captured by moving RS cameras produce distortions known as the RS effect [17], which defeats vital steps (*e.g.* absolute [1] and relative [4] pose estimation) in SfM and SLAM, including BA [2, 9, 11, 12, 14]. Hence, handling the RS effect in BA is crucial for 3D computer vision applications.

1.1. Related Work

Video-based Methods. Hedborg *et al.* [10] use an RS video sequence to solve RSSfM and present an RSBA algorithm under the smooth motion assumption between consecutive frames in [9]. Similarly, Im *et al.* [11] propose a small motion interpolation-based RSBA algorithm specifically for narrow-baseline RS image sequences. Zhuang *et al.* [24] further address this setting by presenting 8pt and 9pt linear solvers, which are developed to recover the relative pose of

* Authors contributed equally
† Corresponding author: yizhenlao@hnu.edu.cn
Project page: https://delinqu.github.io/NW-RSBA

an RS camera that undergoes constant velocity and acceleration motion, respectively. Differently, a spline-based trajectory model to better reformulate the RS camera motion between consecutive frames is proposed by [19].

Unordered RSBA. An unordered image set is the standard input for SfM. Albl *et al.* [2] address the unordered RSSfM problem and point out the planar degeneracy configuration of RSSfM. Ito *et al.* [12] attempt to solve RSSfM by establishing an equivalence with self-calibrated SfM based on the pure rotation instantaneous-motion model and affine camera assumption, while the work of [15] draws the equivalence between RSSfM and non-rigid SfM. A camera-based RSBA has been proposed in [14] to simulate the actual camera projection which has the resilience ability against planar degeneracy.

Direct Methods. Unlike the feature-based BA methods that minimize reprojection errors of keypoints, photometric-based BA methods maximize the photometric consistency among each pair of frames instead (*e.g.* [6, 7]). To handle the RS effect, the works of [13] and [20] present direct semi-dense and direct spare SLAM pipelines, respectively.

1.2. Motivations

Although existing works of RSBA have shown promising results, we argue that they generally lead to either highly complex constraints on inputs, overly restrictive motion models, or time-consuming which limit application fields.

More general: *1)* Video-based [9, 11, 19, 24] and direct methods [13, 20] that use video sequence as input are often not desirable to be processed frame by frame which results in high computational power requirements. *2)* The unordered image set is the typical input for classical SfM pipeline (*e.g.* [23]). *3)* Even the BA step in current popular SLAM systems (*e.g.* [18]) only optimizes keyframes instead of all the consecutive frames.

More effective: [2] points out that when the input images are taken with similar readout directions, RSBA fails to recover structure and motion. The proposed solution is simply to avoid the degeneracy configurations by taking images with perpendicular readout directions. This solution considerably limits the use in common scenarios.

More efficient: GSBA argumentation with the RS motion model has been reported as time-consuming, except for the work of [9] to accelerate video-based RSBA. However, the acceleration of unordered RSBA has never been addressed in the existing works [2, 12, 14, 15].

In summary, an accurate and fast solution to unordered images RSBA without camera motion assumptions, readout direction is still missing. Such a method would be a vital step in the potential widespread deployment of 3D vision with RS imaging systems.

1.3. Contribution

In this paper, we present a novel RSBA solution and tackle the challenges that remained in the previous works. To this end, *1)* we investigate the influence of normalization to the image point on RSBA performance and show its advantages. *2)* Then we present an analytical model for the visual residual covariance, which can standardize the reprojection error during BA, consequently improving the overall accuracy. *3)* Moreover, the combination of **N**ormalization and covariance standardization **W**eighting in **RSBA** (*NW-RSBA*) can avoid common planar degeneracy without constraining the capture manner. *4)* Besides, we propose its acceleration strategy based on the sparsity of the Jacobian matrix and Schur complement. As shown in Fig. 1 that *NW-RSBA* can be easily implemented and plugin GSSfM and GSSLAM as RSSfM and RSSLAM solutions.

In summary, the main contributions of this paper are:

- We first thoroughly investigate image point normalization's influence on RSBA and propose a probability-based weighting algorithm in the cost function to improve RSBA performance. We apply these two insights in the proposed RSBA framework and demonstrate its acceleration strategy.
- The proposed RSBA solution *NW-RSBA* can easily plugin multiple existing GSSfM and GSSLAM solutions to handle the RS effect. The experiments show that *NW-RSBA* provides more accurate results and $10\times$ speedup over the existing works. Besides, it avoids planar degeneracy with the usual capture manner.

2. Formulation of RSBA

We formulate the problem of RSBA and provide a brief description of three RSBA methods in existing works. Since this section does not contain our contributions, we give only the necessary details to follow the paper.

• **Direct-measurement-based RS model:** Assuming a 3D point $\mathbf{P}_i = [X_i \ Y_i \ Z_i]$ is observed by a RS camera j represented by measurement $\mathbf{m}_i^j$ in the image domain. The projection from 3D world to the image can be defined as:

$$\mathbf{m}_i^j = \begin{bmatrix} u_i^j & v_i^j \end{bmatrix}^\top = \Pi(\mathbf{K}\mathbf{P}_i^{cj}), \tag{1}$$

$$\mathbf{P}_i^{cj} = [X_i^{cj} \ Y_i^{cj} \ Z_i^{cj}]^\top = \mathbf{R}^j(v_i^j)\mathbf{P}_i + \mathbf{t}^j(v_i^j), \tag{2}$$

where $\Pi([X \ Y \ Z]^\top) = \frac{1}{Z}[X \ Y]^\top$ is the perspective division and $\mathbf{K}$ is the camera intrinsic matrix [8]. $\mathbf{R}^j(v_i^j) \in \mathbf{SO}(3)$ and $\mathbf{t}^j(v_i^j) \in \mathbb{R}^3$ define the camera rotation and translation respectively when the row index of measurement v_i^j is acquired. Assuming constant camera motion during frame capture, we can model the instantaneous-motion as:

$$\mathbf{R}^j(v_i^j) = (\mathbf{I} + [\boldsymbol{\omega}^j]_\times v_i^j)\mathbf{R}_0^j, \qquad \mathbf{t}^j(v_i^j) = \mathbf{t}_0^j + \mathbf{d}^j v_i^j, \tag{3}$$

where $[\boldsymbol{\omega}^j]_\times$ represents the skew-symmetric matrix of vector $\boldsymbol{\omega}^j$ and $\mathbf{t}_0^j$, $\mathbf{R}_0^j$ is the translation and rotation matrix when the first row is observed. While $\mathbf{d}^j = [d_x^j, d_y^j, d_z^j]^\top$ is the linear velocity vector and $\boldsymbol{\omega}^j = [\omega_x^j, \omega_y^j, \omega_z^j]^\top$ is the angular velocity vector. Such model was adopted by [2, 4, 15]. Notice that RS instantaneous-motion is a function of v_i^j, which we named the direct-measurement-based RS model.

• **Normalized-measurement-based RS model:** By assuming a pre-calibrated camera, one can transform an image measurement $\mathbf{m}_i^j$ with $\mathbf{K}$ to the normalized measurement $[\mathbf{q}_i^{j\top}, 1]^\top = \mathbf{K}^{-1}[\mathbf{m}_i^{j\top}, 1]^\top$. Thus, the projection model and camera instantaneous-motion become:

$$\mathbf{q}_i^j = \begin{bmatrix} c_i^j & r_i^j \end{bmatrix}^\top = \Pi(\mathbf{R}^j(r_i^j)\mathbf{P}_i + \mathbf{t}^j(r_i^j)), \tag{4}$$

$$\mathbf{R}^j(r_i^j) = (\mathbf{I} + [\boldsymbol{\omega}^j]_\times r_i^j)\mathbf{R}_0^j, \quad \mathbf{t}^j(r_i^j) = \mathbf{t}_0^j + \mathbf{d}^j r_i^j. \tag{5}$$

In contrast to the direct-measurement-based RS model, $\mathbf{t}_0^j$ and $\mathbf{R}_0^j$ are the translation and rotation when the optical centre row is observed. $\boldsymbol{\omega}^j$, $\mathbf{d}^j$ are scaled by camera focal length. We name such model the normalized-measurement-based RS model, which was used in [2, 9, 12, 24].

• **Rolling Shutter Bundle Adjustment:** The non-linear least squares optimizers are used to find a solution $\boldsymbol{\theta}^*$ including camera poses $\mathbf{R}^*, \mathbf{t}^*$, instantaneous-motion $\boldsymbol{\omega}^*, \mathbf{d}^*$ and 3D points $\mathbf{P}^*$ by minimizing the reprojection error $\mathbf{e}_i^j$ from point i to camera j over all the camera index in set $\mathcal{F}$ and corresponding 3D points index in subset $\mathcal{P}_j$:

$$\boldsymbol{\theta}^* = \{\mathbf{P}^*, \mathbf{R}^*, \mathbf{t}^*, \boldsymbol{\omega}^*, \mathbf{d}^*\} = \arg\min_{\boldsymbol{\theta}} \sum_{j \in \mathcal{F}} \sum_{i \in \mathcal{P}_j} \left\| \mathbf{e}_i^j \right\|_2^2. \tag{6}$$

(1) **Direct-measurement-based RSBA:** [5] uses the direct-measurement-based RS model and compute the reprojection error as: $\mathbf{e}_i^j = \mathbf{m}_i^j - \Pi(\mathbf{K}(\mathbf{R}^j(v_i^j)\mathbf{P}_i + \mathbf{t}^j(v_i^j)))$. We name this strategy *DM-RSBA*.

(2) **Normalized-measurement-based RSBA:** [2] uses the normalized-measurement-based RS model and compute the reprojection error as: $\mathbf{e}_i^j = \mathbf{q}_i^j - \Pi(\mathbf{R}^j(r_i^j)\mathbf{P}_i + \mathbf{t}^j(r_i^j))$. We name this strategy *NM-RSBA*.

(3) **Direct-camera-base RSBA:** Lao *et al.* [14] argue both *DM-RSBA* and *NM-RSBA* can not simulate the actual projection. So [14] proposes a camera-based approach that uses camera pose and instantaneous motion to compute the reprojection without using either v_i^j or r_i^j. We name this strategy *DC-RSBA*.

3. Methodology

In this section, we present a novel RSBA algorithm called normalized weighted RSBA (*NW-RSBA*). The main idea is to use measurement normalization (section 3.1) and covariance standardization weighting (section 3.2) jointly during the optimization. Besides, we provide a two-stage Schur complement strategy to accelerate *NW-RSBA* in section 3.3. The pipeline of *NW-RSBA* is shown in Alg. 1. The main differences between *NW-RSBA* and existing methods [2, 5, 14] are summarized in Tab. 1.

Algorithm 1: Normalized Weighted RSBA

Input : Initial rolling shutter camera poses and points as state vector $\boldsymbol{\theta}$. Observed point measurement in normalized image coordinate.

Output: Refined state vector $\boldsymbol{\theta}^*$

```
while (not reach max iteration) and (not satisfy stopping criteria) do
    for Each camera j ∈ F do
        for Each point i ∈ P_j do
            Calculate error ê_i^j using Eq.(13);
            Construct matrix J_i^j (supplemental material);
            Parallel connect J_i^j to J;
            Stack ê_i^j into ê;
        end
    end
    Solve equation J^T J δ = −J^T ê using Alg. 2;
    Update state vector θ using δ;
end
```

Table 1. Comparison of the proposed *NW-RSBA* and existing general unordered RSBA methods [2, 5, 14]. D: Direct measurement; N: normalized measurement; M-based: measurement-based projection model; C-based: camera-based projection model.

	DM-RSBA [5]	*NM-RSBA* [2]	*DC-RSBA* [14]	*NW-RSBA*
Normalization	D	N	D	N
Reprojecting computation	M-based	M-based	C-based	M-based
Analytical Jacobian	×	√	×	√
BA acceleration	×	×	×	√

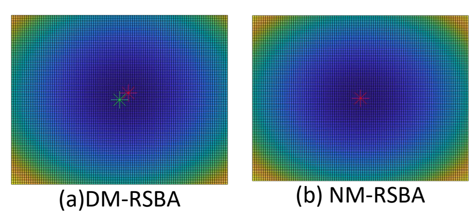

Figure 2. Reprojection loss surface of (a) *DM-RSBA* [5] and (b) *NM-RSBA* [2] under the same configuration. The x-axis and y-axis are d_x^j and d_y^j of camera motion, respectively. The * are the ground truth solutions while the * are the minimums in loss surfaces.

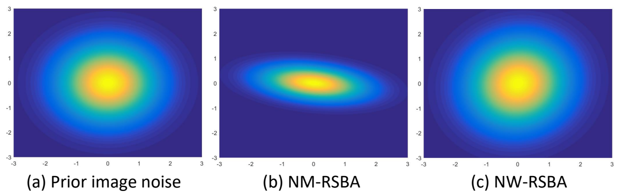

Figure 3. Simulated distributions of (a) prior Gaussian image noise, error term of (b) *NM-RSBA* [2], and the proposed (c) *NW-RSBA*.

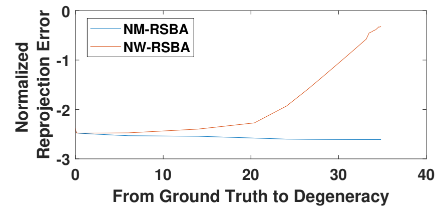

Figure 4. The normalized reprojection error comparison between our proposed method *NW-RSBA* and *NM-RSBA* [2] along the degeneracy process.

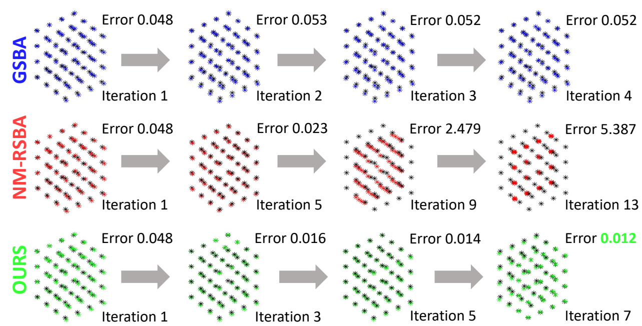

Figure 5. Analysis of the degeneracy in synthetic cube scene captured by five cameras with parallel readout direction. The reconstruction of *NM-RSBA* [2] (red) collapses to the plane gradually while *NW-RSBA* (green) provides the most accurate result.

3.1. Measurement Normalization

To the best of our knowledge, no prior research has investigated the influence of normalization on RSBA performance. Thus, we conduct a synthetic experiment to evaluate the impact of normalization by comparing the performances of *DM-RSBA* [5] and *NM-RSBA* [2]. The results in Fig. 8 show that the normalization significantly improves the RSBA accuracy. The improvement comes from the more precise instantaneous-motion approximation of low-order Taylor expansion in *NM-RSBA*. In *DM-RSBA*, the errors on the image plane induced by the approximate have the same directions and grow exponentially with the increase of the row index. Thus, the optimizer will shift the solution away from the ground truth to equally average the error among all observations. In contrast, the error distribution in *NM-RSBA* is inherently symmetrical due to the opposite direction with the row varying from the optical center, thus exhibiting significantly lower bias. A synthetic example shown in Fig. 2 verifies our insight that *NM-RSBA* has an unbiased local minimum over *DM-RSBA*.

3.2. Normalized Weighted RSBA

Based on the measurement normalization, we further present a normalized weighted RSBA (*NW-RSBA*) by modelling image noise covariance.

• **Weighting the reprojection error:** In contrast to existing works [2, 5, 14], we consider the image noise in RSBA by weighting the squared reprojection error with the inverse covariance matrix of the error, which is also known as standardization of the error terms [22]. Thus the cost function in Eq. (6) becomes:

$$\boldsymbol{\theta}^{*}=\underset{\boldsymbol{\theta}}{\arg\min}\sum_{j\in\mathcal{F}}\sum_{i\in\mathcal{P}_j}{\mathbf{e}_i^j}^{\top}{\boldsymbol{\Sigma}_i^j}^{-1}\mathbf{e}_i^j, \tag{7}$$

where $\mathbf{e}_i^j$ is the reprojection error computed by normalized-measurement-based approach described in Eq. (4). By assuming the image measurement $\mathbf{q}_i^j$ follows a prior Gaussian noise: $\mathbf{n}_i^j\sim\mathcal{N}(0,\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^{\top})$, the newly introduced covariance matrix of the error $\boldsymbol{\Sigma}_i^j$ is defined as follows (proof in the supplemental material):

$$\boldsymbol{\Sigma}_i^j=\mathbf{C}_i^j\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^{\top}{\mathbf{C}_i^j}^{\top}, \tag{8}$$

$$\mathbf{C}_i^j=\begin{bmatrix}1&0\\0&1\end{bmatrix}-\begin{bmatrix}\frac{1}{Z^{c_i^j}}&0\\0&\frac{1}{Z^{c_i^j}}\\\frac{-X^{c_i^j}}{{Z^{c_i^j}}^2}&\frac{-Y^{c_i^j}}{{Z^{c_i^j}}^2}\end{bmatrix}^{\top}([\boldsymbol{\omega}^j]_{\times}\mathbf{R}^j\mathbf{P}_i+\mathbf{d}^j)\begin{bmatrix}0\\1\end{bmatrix}^{\top}, \tag{9}$$

$$\mathbf{W}=\begin{bmatrix}1/f_x&0\\0&1/f_y\end{bmatrix}, \tag{10}$$

where $\mathbf{C}_i^j$ and $\mathbf{W}$ are 2×2 auxiliary matrices. f_x and f_y are focal lengths.

• **Advantages of noise covariance weighting:** Note that the standardisation in Eq. (7) scaling every term by its inverse covariance matrix ${\boldsymbol{\Sigma}_i^j}^{-1}$, so that all terms end up with isotropic covariance [22]. In Fig. 3, we visualize the influence of error terms standardization, which re-scales the prior covariance ellipse back to a unit one. We interpret the re-weighting standardization leads two advantages:

(1) **More accurate:** With the re-weighting standardization in Eq. (7), features with a high variance which means a high probability of a large reprojection error, are offered less confidence to down-weighting their influence on the total error. Synthetic experiments in section 4.1 demonstrate it outperforms *NM-RSBA* under various noise levels.

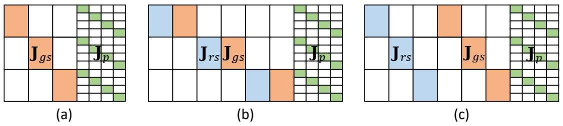

Figure 6. Example Jacobian matrices with 4 points and 3 cameras in (a) GSBA, RSBA with (b) series and (c) parallel connection.

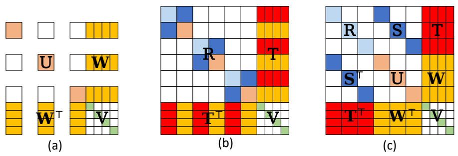

Figure 7. Example Hessian matrices with 4 points and 3 cameras in (a) GSBA, RSBA with(b) series and (c) parallel connection.

(2) **Handle planar degeneracy:** Though the proposed *NW-RSBA* uses measurement-based projection during degeneracy, it still provides stable BA results and even outperforms *C-RSBA* with the noise covariance weighting. As demonstrated in [2], under the planar degeneracy configuration, the y-component of the reprojection error will reduce to zero, which denotes that the covariance matrix holds a zero element in the y direction. *NW-RSBA* explicitly modeled the error covariance and standardized it to isotropy. Thus, the proposed method exponentially amplifies the error during degeneracy, as shown in Fig. 4, and prevents the continuation of the decline from ground truth to the degenerated solution (proofs can be found in the supplemental material). A synthetic experiment shown in Fig. 5 verifies that *NM-RSBA* easily collapses into degeneracy solutions while *NW-RSBA* provide stable result and outperforms the *GSBA*.

• **Jacobian of *NW-RSBA*:** For more convenient optimization, we reformulate covariance matrix $\mathbf{\Sigma}_i^j$ as a standard least square problem using decomposition:

$$\mathbf{\Sigma}_i^{j\,-1} = \mathbf{C}_i^{j\,-\top}\mathbf{W}^{-\top}\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{W}^{-1}\mathbf{C}_i^{j\,-1}. \tag{11}$$

By substituting Eq. (11) into (7), we get a new cost function:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \sum_{j\in\mathcal{F}}\sum_{i\in\mathcal{P}_j} \hat{\mathbf{e}}_i^{j\top}\hat{\mathbf{e}}_i^j, \tag{12}$$

$$\text{where, } \hat{\mathbf{e}}_i^j = \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{W}^{-1}\mathbf{C}_i^{j\,-1}\mathbf{e}_i^j. \tag{13}$$

We derive the analytical Jacobian matrix of $\hat{\mathbf{e}}_i^j$ in Eq. (12) using the chain rule in the supplemental material.

3.3. NW-RSBA Acceleration

Based on the sparsity of the Jacobian, the marginalization [16, 22] with Schur complement has achieved significant success in accelerating GSBA. However, the acceleration strategy has never been addressed for the general unordered RSBA in [2, 5, 14]. As shown in Fig. 6(b)(c), we can organize the RSBA Jacobian in two styles:

(1) **Series connection:** By connecting camera pose and instantaneous-motion in the Jacobian matrix (Fig. 7(b)) as an entirety, we can use the **one-stage Schur complement** technique [16] to marginalize out the 3D point and compute the update state vector for $\mathbf{R}, \mathbf{t}, \boldsymbol{\omega}, \mathbf{d}$ first, followed by back substitution for update state vector of points $\mathbf{P}$.

(2) **Parallel connection:** Due to the independence between camera pose and instantaneous-motion in the Jacobian matrix (Fig. 7(c)), we propose a **two-stage Schur complement** strategy to accelerate RSBA. When solving the non-linear least square problem (*e.g.* Gauss-Newton), the approximate Hessian matrix for Eq. (12) is defined as

$$\mathbf{J}^\top\mathbf{J} = \begin{bmatrix} \mathbf{J}_{rs}{}^\top\mathbf{J}_{rs} & \mathbf{J}_{rs}{}^\top\mathbf{J}_{gs} & \mathbf{J}_{rs}{}^\top\mathbf{J}_{p} \\ \mathbf{J}_{gs}{}^\top\mathbf{J}_{rs} & \mathbf{J}_{gs}{}^\top\mathbf{J}_{gs} & \mathbf{J}_{gs}{}^\top\mathbf{J}_{p} \\ \mathbf{J}_{p}{}^\top\mathbf{J}_{rs} & \mathbf{J}_{p}{}^\top\mathbf{J}_{gs} & \mathbf{J}_{p}{}^\top\mathbf{J}_{p} \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{S} & \mathbf{T} \\ \mathbf{S}^\top & \mathbf{U} & \mathbf{W} \\ \mathbf{T}^\top & \mathbf{W}^\top & \mathbf{V} \end{bmatrix}, \tag{14}$$

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{rs} & \mathbf{J}_{gs} & \mathbf{J}_p \end{bmatrix} = \begin{bmatrix} \frac{\partial\hat{\mathbf{e}}}{\partial\mathbf{x}_{rs}} & \frac{\partial\hat{\mathbf{e}}}{\partial\mathbf{x}_{gs}} & \frac{\partial\hat{\mathbf{e}}}{\partial\mathbf{x}_{p}} \end{bmatrix}, \tag{14a}$$

$$\hat{\mathbf{e}} = \left[\left\{\hat{\mathbf{e}}_i^j\right\}\right], \quad \mathbf{x}_{gs} = \left[\left\{\mathbf{R}^j\right\} \quad \left\{\mathbf{t}^j\right\}\right], \tag{14b-1}$$

$$\mathbf{x}_{rs} = \left[\left\{\boldsymbol{\omega}^j\right\} \quad \left\{\mathbf{d}^j\right\}\right], \quad \mathbf{x}_p = \left[\left\{\mathbf{P}_i\right\}\right], \tag{14b-2}$$

where $\mathbf{R}$, $\mathbf{U}$, $\mathbf{V}$, $\mathbf{S}$, $\mathbf{T}$ and $\mathbf{W}$ are submatrices computed by the derivations $\mathbf{J}_{rs}$, $\mathbf{J}_{gs}$ and $\mathbf{J}_p$. As Alg. 2 shown that the two-stage Schur complement strategy consists of 3 steps:

▷ **Step 1: Construct normal equation.** In each iteration, using this form of the Jacobian and corresponding state vectors and the error vector, the normal equation follows

$$\mathbf{J}^\top\mathbf{J}\boldsymbol{\delta} = \begin{bmatrix} \mathbf{R} & \mathbf{S} & \mathbf{T} \\ \mathbf{S}^\top & \mathbf{U} & \mathbf{W} \\ \mathbf{T}^\top & \mathbf{W}^\top & \mathbf{V} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_{rs} \\ \boldsymbol{\delta}_{gs} \\ \boldsymbol{\delta}_{p} \end{bmatrix} = -\begin{bmatrix} \mathbf{t} \\ \mathbf{u} \\ \mathbf{v} \end{bmatrix} = -\mathbf{J}^\top\hat{\mathbf{e}}, \tag{15}$$

Algorithm 2: Solve the normal equation using two-stage Schur complement

Input : Jacobian matrix $\mathbf{J}$ and weighted error vector $\hat{\mathbf{e}}$

Output: Update state vector $\boldsymbol{\delta}$

Compute Schur complement matrix $\mathbf{S}_p$ and $\mathbf{S}_{rs}$ using Eq. (16) and (17);

Compute auxiliary vectors $\mathbf{t}^*$ and $\mathbf{u}^*$using Eq. (19) and (18);

Solve Eq. (20) cascadingly:

- Get $\boldsymbol{\delta}_{rs}$ by solving $\mathbf{S}_{rs}\boldsymbol{\delta}_{rs} = -\mathbf{t}^*$;
- Get $\boldsymbol{\delta}_{gs}$ by solving $\mathbf{U}^*\boldsymbol{\delta}_{gs} = -\mathbf{u}^* - \mathbf{S}^{*\top}\boldsymbol{\delta}_{rs}$;
- Get $\boldsymbol{\delta}_{p}$ by solving $\mathbf{V}\boldsymbol{\delta}_{p} = -\mathbf{v} - \mathbf{T}^\top\boldsymbol{\delta}_{rs} - \mathbf{W}^\top\boldsymbol{\delta}_{gs}$;
- Stack $\boldsymbol{\delta}_{gs}$, $\boldsymbol{\delta}_{rs}$, $\boldsymbol{\delta}_{p}$ into $\boldsymbol{\delta}$;

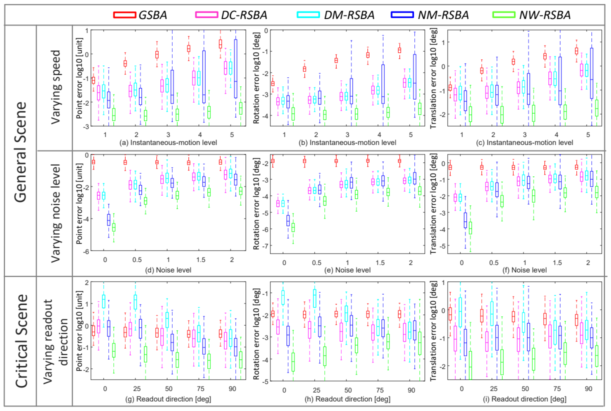

Figure 8. Camera pose (2nd and 3rd columns) and reconstruction (1st column) errors of *GSBA*, *DC-RSBA*, *DM-RSBA*, *NM-RSBA* and *NW-RSBA* with increasing angular and linear velocity (1st row) and noise levels in the image (2nd row) in a general scene, also with increasing readout directions in a degeneracy scene (3rd row).

where $\boldsymbol{\delta}_{gs}$, $\boldsymbol{\delta}_{rs}$ and $\boldsymbol{\delta}_p$ are the update state vectors to $\mathbf{x}_{gs}$, $\mathbf{x}_{rs}$, and $\mathbf{x}_p$, while $\mathbf{t}$, $\mathbf{u}$, and $\mathbf{v}$ are the corresponding descent direction. Such formed normal equations show a block sparsity, suggesting that we can efficiently solve it.

▷ **Step 2: Construct Schur complement.** We construct two-stage Schur complements $\mathbf{S}_p$, $\mathbf{S}_{rs}$ and two auxiliary vectors $\mathbf{u}^*$ and $\mathbf{t}^*$ to Eq. (15) as

$$\mathbf{S}_p = \begin{bmatrix} \mathbf{R}^* & \mathbf{S}^* \\ \mathbf{S}^{*\top} & \mathbf{U}^* \end{bmatrix} = \begin{bmatrix} \mathbf{R} - \mathbf{T}\mathbf{V}^{-1}\mathbf{T}^\top & \mathbf{S} - \mathbf{T}\mathbf{V}^{-1}\mathbf{W}^\top \\ \mathbf{S}^\top - \mathbf{W}\mathbf{V}^{-1}\mathbf{T}^\top & \mathbf{U} - \mathbf{W}\mathbf{V}^{-1}\mathbf{W}^\top \end{bmatrix}, \quad (16)$$

$$\mathbf{S}_{rs} = \mathbf{R}^* - \mathbf{S}^*{\mathbf{U}^*}^{-1}\mathbf{S}^{*\top}, \quad (17)$$

$$\mathbf{u}^* = \mathbf{u} - \mathbf{W}\mathbf{V}^{-1}\mathbf{v}, \quad (18)$$

$$\mathbf{t}^* = \mathbf{t} - \mathbf{T}\mathbf{V}^{-1}\mathbf{v} - \mathbf{S}^*\mathbf{U}^{*-1}\mathbf{u}^*. \quad (19)$$

▷ **Step 3: Orderly solve $\boldsymbol{\delta}_{gs}$, $\boldsymbol{\delta}_{rs}$ and $\boldsymbol{\delta}_p$.** Based on $\mathbf{S}_p$, $\mathbf{S}_{rs}$, $\mathbf{u}^*$ and $\mathbf{t}^*$, we reformulate the normal equation as

$$\begin{bmatrix} \mathbf{S}_{rs} & \mathbf{0} & \mathbf{0} \\ \mathbf{S}^{*\top} & \mathbf{U}^* & \mathbf{0} \\ \mathbf{T}^\top & \mathbf{W}^\top & \mathbf{V} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_{rs} \\ \boldsymbol{\delta}_{gs} \\ \boldsymbol{\delta}_p \end{bmatrix} = - \begin{bmatrix} \mathbf{t}^* \\ \mathbf{u}^* \\ \mathbf{v} \end{bmatrix}, \quad (20)$$

which enables us to compute $\boldsymbol{\delta}_{rs}$ first, and then back substitutes the results to get $\boldsymbol{\delta}_{gs}$. Finally, we can obtain $\boldsymbol{\delta}_p$ based on the 3rd row of Eq. (20).

3.4. Implementation

We follow Alg. 1 to implement the proposed *NW-RSBA* in C++. The implemented *NW-RSBA* can serve as a little module and can be easily plug-in such context:

- **RS-SfM:** We augment VisualSFM [23] by shifting the incremental GSBA pipeline with the proposed *NW-RSBA*.
- **RS-SLAM:** We augment Orb-SLAM [18] by replacing the local BA and full BA modules with *NW-RSBA*.

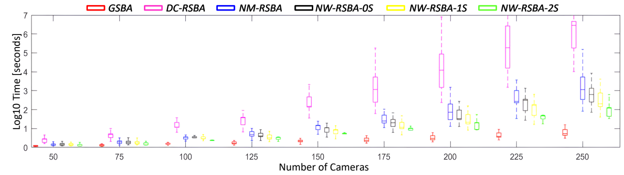

Figure 9. Time cost of *GSBA* [16], *DC-RSBA* [14], *NM-RSBA* [2], *NW-RSBA-0S* (without Schur complement), *NW-RSBA-1S* (one-stage Schur complement to Jacobian matrices with series connection), and proposed *NW-RSBA-2S* (two-stage Schur complement to Jacobian matrices with parallel connection) with increasing camera number and fixed point number.

4. Experimental Evaluation

In our experiments, the proposed method is compared to three state-of-the-art unordered RSBA solutions:*1) GSBA*: SBA [16]. *2) DC-RSBA*: direct camera-based RSBA [14]. *3) DM-RSBA*: direct measurement-based RSBA [5]. *4) NM-RSBA*: normalized-measurement-based RSBA [2]. *5) NW-RSBA*: proposed normalized weighted RSBA.

4.1. Synthetic Data

Settings and metrics. We simulate 5 RS cameras located randomly on a sphere pointing at a cubical scene. We compare all methods by varying the speed, the image noise, and the readout direction. The results are obtained after collecting the errors over 300 trials per epoch. We measure the reconstruction errors and pose errors.

Results. *1) **Varying Speed.*** The results in Fig. 8(a)(b)(c) show that the estimated errors of *GSBA* grow with speed while *DC-RSBA*, *DM-RSBA* and *NM-RSBA* achieve better results with slow kinematics. The proposed *NW-RSBA* provides the best results under all configurations. *2) **Varying Noise Level.*** In Fig. 8(d)(e)(f), *GSBA* shows better robustness to noise but with lower accuracy than RS methods. The proposed *NW-RSBA* achieves the best performance with all noise levels. *3) **Varying Readout Direction.*** We evaluate five methods with varying readout directions of the cameras by increasing the angle from parallel to perpendicular. Fig. 8(g)(h)(i) show that under a small angle, the reconstruction error of *DM-RSBA*, *DC-RSBA* and *DM-RSBA* grow dramatically even bigger than *GSBA*, suggesting a degenerated solution. In contrast, *NW-RSBA* provides stable results under all settings, even with the parallel readout direction.

Runtime. As shown in Fig. 9 that without analytical Jacobian, *DC-RSBA* is the slowest one while the proposed *NW-RSBA* achieves similar efficiency as *NM-RSBA*. However, by using acceleration strategies, *NW-RSBA-1S* and *NW-RSBA-2S* reduce the overall runtime. Note that *NW-RSBA-2S* achieves an order of magnitude faster than *NM-RSBA*.

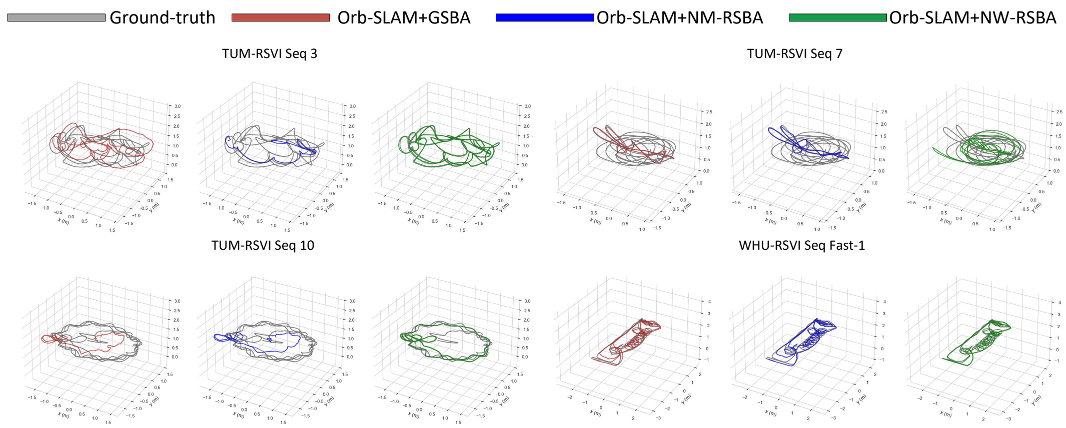

Figure 10. Ground truth and trajectories estimated by *GSBA* [16], *NM-RSBA* [2] and proposed *NW-RSBA* after Sim(3) alignment on 3 sequences from TUM-RSVI [21] and 1 sequence from WHU-RSVI [3] datasets.

Table 2. Statistics of 10 sequences from TUM-RSVI [21] and 2 sequences from WHU-RSVI [3] datasets. The realtime factor ϵ and tracking duration $\underline{DUR}$ of *Orb-SLAM*, *Orb-SALM+NM-RSBA* [2], and proposed *Orb-SLAM+NW-RSBA*.

Seq	Duration [s]	length [m]	Realtime factor $\epsilon\uparrow$ | Tracking duration $\underline{DUR}\uparrow$		
			Orb-SLAM [18]	Orb-SLAM +NM-RSBA [2]	Orb-SLAM +NW-RSBA
#1	40	46	1.47 | 0.50	1.48 | 0.28	1.38 | **1**
#2	27	37	1.51 | 0.90	1.40 | 0.81	1.40 | **1**
#3	50	44	1.47 | 0.58	1.41 | 0.36	1.39 | **1**
#4	38	30	1.61 | 1	1.35 | 1	1.56 | 1
#5	85	57	1.51 | 1	1.28 | 1	1.38 | 1
#6	43	51	1.47 | 0.76	1.37 | 0.76	1.38 | **1**
#7	39	45	1.61 | 0.89	1.47 | 0.97	1.49 | **1**
#8	53	46	1.56 | 0.79	1.37 | 0.96	1.35 | **1**
#9	45	46	1.61 | 0.14	1.51 | 0.23	1.55 | **0.42**
#10	54	41	1.56 | 0.29	1.46 | 0.29	1.47 | **1**
t1-fast	28	50	1.92 | 1	1.51 | 1	1.81 | 1
t2-fast	29	53	1.92 | 1	1.40 | 1	1.67 | 1

4.2. Real Data

Datasets and metrics. We compare all the RSBA methods in two publicly available RS datasets: WHU-RSVI [3] dataset[1], TUM-RSVI [21] dataset[2]. In this section, we use three evaluation metrics, namely ATE e_{ate} (absolute trajectory error) [21], tracking duration $\underline{DUR}$ (the ratio of the successfully tracked frames out of the total frames) and realtime factor ϵ (sequence's actual duration divided by the algorithm's processing time).

4.2.1 RSSLAM

We compare the performance of conventional GS-based *Orb-SLAM* [18] versus augmented versions with *NM-RSBA* [2] and proposed *NW-RSBA* on 12 RS sequences.

Real-time factor and tracking duration. Tab. 2 shows the statistics about 12 RS sequences and the performance of three approaches run on an AMD Ryzen 7 CPU. The results verify that all three methods achieve real-time performance. One can also confirm that the proposed *Orb-SLAM+NW-RSBA* is slower than *Orb-SLAM* by a factor roughly around 1.2 but is slightly faster than *Orb-SLAM+NM-RSBA*. As for tracking duration, *Orb-SLAM* and *Orb-SLAM+NM-RSBA* [2] fail with an average $\underline{DUR} < 0.5$ in most sequences once the camera moves aggressively enough. In contrast, the proposed *Orb-SLAM+NW-RSBA* achieves completed tracking with $\underline{DUR} = 1$ in almost all sequences.

Absolute trajectory error. The **ATE** results on WHU-RSVI and TUM-RSVI datasets demonstrate that the proposed *Orb-SLAM+NW-RSBA* is superior to *Orb-SLAM* and *Orb-SLAM+NM-RSBA* when dealing with RS effect. Qualitatively, this is clearly visible in Figs. 1 and 10. The sparse 3D reconstructions look much cleaner for *Orb-SLAM+NW-RSBA* and close to the ground truth. The quantitative difference also becomes apparent in Tab. 3. *Orb-SLAM+NW-RSBA* outperforms *Orb-SLAM* and *Orb-SLAM+NM-RSBA* both in terms of accuracy and stability.

4.2.2 RSSfM

Quantitative Ablation Study. We randomly choose 8 frames from each of the 12 RS sequences to generate 12 unordered SfM datasets and evaluate the RSBA performance via average ATE and runtime. Besides, we ablate the *NW-RSBA* and compare quantitatively with related approaches. The results are presented in Tab. 4. The baseline methods' performance show that *NW-RSBA* obtains ATE of 0.007, which is half of the second best method *DC-RSBA* and nearly 3% of *GSBA*. The removal of normalization from proposed *NW-RSBA* adversely increases ATE up to 100%. The removal of covariance weighting from *NW-RSBA* adversely impacts the camera pose estimation quality with ATE growth from 0.007 to 0.020. We believe that covariance weighting helps BA leverage the RS effect and random

[1] http://aric.whu.edu.cn/caolike/2019/11/05/the-whu-rsvi-dataset
[2] https://vision.in.tum.de/data/datasets/rolling-shutter-dataset

Table 3. Absolute trajectory error (ATE) of different RSBA methods after Sim(3) alignment to ground truth. The best results are shown in green. Since some methods will lose tracking without processing the whole sequence, thus we highlight the background of each cell with different colours depending on its corresponding *DUR* value. Specifically, *DUR* > 0.9, $0.5 <$ *DUR* $\leqslant 0.9$ and *DUR* $\leqslant 0.5$ are highlighted in light green , cyan , and orange .

Input	Methods	ATE↓											
		TUM-RSVI [21]										WHU-RSVI [3]	
		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	t1-fast	t2-fast
GS data	Orb-SLAM [18]	0.015	0.013	0.018	0.107	0.030	0.013	0.054	0.053	0.020	0.024	0.044	0.008
RS data	Orb-SLAM [18]	0.059	0.100	0.411	0.126	0.055	0.044	0.217	0.218	0.176	0.373	0.237	0.018
	Orb-SLAM+NM-RSBA [2]	0.115	0.088	0.348	0.120	0.062	0.060	0.251	0.246	0.156	0.307	0.204	0.030
	Orb-SLAM+NW-RSBA (ours)	0.011	0.008	0.031	0.071	0.034	0.008	0.260	0.115	0.028	0.108	0.054	0.012

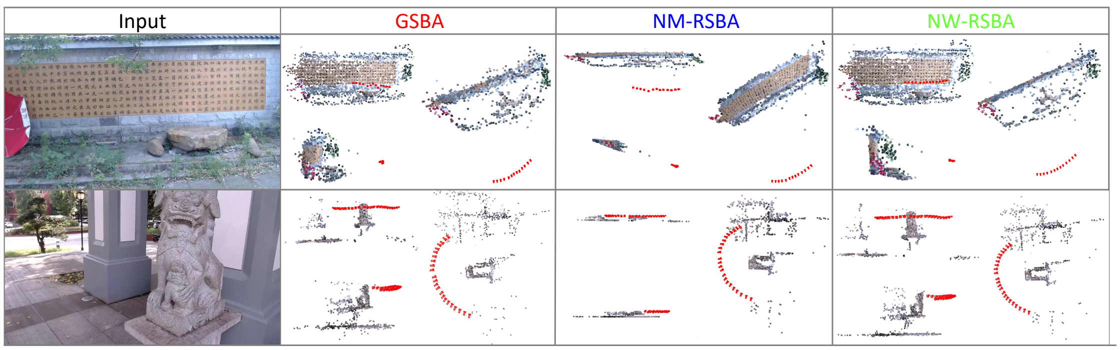

Figure 11. Three-view graph of reconstructions using SfM pipeline with *GSBA* [16], *NM-RSBA* [2] and proposed *NW-RSBA*.

Table 4. Quantitative ablation study of RSSfM on TUM-RSVI [21] and WHU-RSVI [3] datasets. ATE: absolute trajectory error of estimated camera pose in meters (m), Runtime: time cost in seconds (s). Best and second best results are shown in green and blue respectively.

Ablation	**Approach**	**ATE** (m)	**Runtime** (s)
	GSBA [16]	0.210	2.9
	DC-RSBA [14]	0.016	1302
No normalization & weighting	DM-RSBA [5]	0.023	740
No weighting	NM-RSBA [2]	0.020	15.8
No normalization	W-RSBA	0.013	16.1
No Schur complement	NW-RSBA-0S		16.9
with 1-stage Schur complement	NW-RSBA-1S	0.007	13.0
Consolidated	NW-RSBA-2S		9.8

image noise better. We ablate the consolidated *NW-RSBA-2S* to *NW-RSBA-0S* by removing the proposed 2-stage Schur complement strategy and compare to *NW-RSBA-1S*. The increases from 9.8s to 13.0s and 16.9s is observed for average runtime. Despite the fact *NW-RSBA-2S* is slower than *GSBA* by a factor of 3, but still 2 times faster than *NM-RSBA*, 2 orders of magnitude faster than *DC-RSBA* and *DM-RSBA*.

Qualitative Samples. We captured two datasets using a smartphone camera and kept the same readout direction, which is a degeneracy configuration in RSBA and will quickly lead to a planar degenerated solution for *NM-RSBA* [2]. As shown in Fig. 11 that *NW-RSBA* works better in motion and 3D scene estimation while *GSBA* [16] obtains a deformed reconstruction. Specifically, the sparse 3D reconstructions and recovered trajectories of *NW-RSBA* look much cleaner and smoother than the ones from *GSBA*. *NM-RSBA* reconstructs 3D scenes which collapse into a plane since the datasets contain only one readout direction. In contrast, *NW-RSBA* provides correct reconstructions. The results also verify our discussion in section 3.2 that error covariance weighting can handle the planar degeneracy.

5. Conclusion

This paper presents a novel RSBA solution without any assumption on camera image manner and type of video input. We explain the importance of conducting normalization and present a weighting technique in RSBA, which leads to normalized weighted RSBA. Extensive experiments in real and synthetic data verify the effectiveness and efficiency of the proposed *NW-RSBA* method.

Acknowledgements. This work is supported by the National Key R&D Program of China (No. 2022ZD0119003), Nature Science Foundation of China (No. 62102145), and Jiangxi Provincial 03 Special Foundation and 5G Program (Grant No. 20224ABC03A05).

References

[1] Cenek Albl, Zuzana Kukelova, Viktor Larsson, and Tomas Pajdla. Rolling shutter camera absolute pose. *PAMI*, 2019. 1

[2] Cenek Albl, Akihiro Sugimoto, and Tomas Pajdla. Degeneracies in rolling shutter sfm. In *ECCV*, 2016. 1, 2, 3, 4, 5, 6, 7, 8

[3] Like Cao, Jie Ling, and Xiaohui Xiao. The whu rolling shutter visual-inertial dataset. *IEEE Access*, 8:50771–50779, 2020. 7, 8

[4] Yuchao Dai, Hongdong Li, and Laurent Kneip. Rolling shutter camera relative pose: generalized epipolar geometry. In *CVPR*, 2016. 1, 3

[5] Gaspard Duchamp, Omar Ait-Aider, Eric Royer, and Jean-Marc Lavest. A rolling shutter compliant method for localisation and reconstruction. In *VISAPP*, 2015. 3, 4, 5, 6, 8

[6] Jakob Engel, Vladlen Koltun, and Daniel Cremers. Direct sparse odometry. *PAMI*, 2017. 2

[7] Jakob Engel, Thomas Schöps, and Daniel Cremers. Lsd-slam: Large-scale direct monocular slam. In *ECCV*, 2014. 2

[8] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 1, 2

[9] Johan Hedborg, Per-Erik Forssen, Michael Felsberg, and Erik Ringaby. Rolling shutter bundle adjustment. In *CVPR*, 2012. 1, 2, 3

[10] Johan Hedborg, Erik Ringaby, Per-Erik Forssén, and Michael Felsberg. Structure and motion estimation from rolling shutter video. In *ICCV Workshops*, 2011. 1

[11] Sunghoon Im, Hyowon Ha, Gyeongmin Choe, Hae-Gon Jeon, Kyungdon Joo, and In So Kweon. Accurate 3d reconstruction from small motion clip for rolling shutter cameras. *PAMI*, 2018. 1, 2

[12] Eisuke Ito and Takayuki Okatani. Self-calibration-based approach to critical motion sequences of rolling-shutter structure from motion. In *CVPR*, 2016. 1, 2, 3

[13] J. H. Kim, C. Cadena, and I. Reid. Direct semi-dense slam for rolling shutter cameras. In *ICRA*, 2016. 2

[14] Yizhen Lao, Omar Ait-Aider, and Helder Araujo. Robustified structure from motion with rolling-shutter camera using straightness constraint. *Pattern Recognition Letters*, 2018. 1, 2, 3, 4, 5, 6, 8

[15] Yizhen Lao, Omar Ait-Aider, and Adrien Bartoli. Solving rolling shutter 3d vision problems using analogies with non-rigidity. *IJCV*, 2021. 2, 3

[16] Manolis IA Lourakis and Antonis A Argyros. Sba: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software (TOMS)*, 36(1):1–30, 2009. 5, 6, 7, 8

[17] M Meingast, C Geyer, and S Sastry. Geometric models of rolling-shutter cameras. In *OMNIVIS*, 2005. 1

[18] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *T-RO*, 2015. 1, 2, 6, 7, 8

[19] Alonso Patron-Perez, Steven Lovegrove, and Gabe Sibley. A spline-based trajectory representation for sensor fusion and rolling shutter cameras. *IJCV*, 2015. 2

[20] David Schubert, Nikolaus Demmel, Vladyslav Usenko, Jörg Stückler, and Daniel Cremers. Direct sparse odometry with rolling shutter. *ECCV*, 2018. 2

[21] David Schubert, Nikolaus Demmel, Lukas von Stumberg, Vladyslav Usenko, and Daniel Cremers. Rolling-shutter modelling for direct visual-inertial odometry. In *IROS*, 2019. 1, 7, 8

[22] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms*, pages 298–372. Springer, 1999. 4, 5

[23] Changchang Wu et al. Visualsfm: A visual structure from motion system. 2011. 2, 6

[24] Bingbing Zhuang, Loong-Fah Cheong, and Gim Hee Lee. Rolling-shutter-aware differential sfm and image rectification. In *ICCV*, 2017. 1, 2, 3