Revisiting Rolling Shutter Bundle Adjustment: Toward Accurate and Fast Solution

Bangyan Liao\textsuperscript{1,2,*} Delin Qu\textsuperscript{1,3,*} Yifei Xue\textsuperscript{4} Huiqing Zhang\textsuperscript{1} Yizhen Lao\textsuperscript{1,†}
\textsuperscript{1}College of Computer Science and Electronic Engineering, Hunan University
\textsuperscript{2}College of Electrical and Information Engineering, Hunan University
\textsuperscript{3}Shanghai AI Laboratory
\textsuperscript{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 Normalization and covariance standardization Weighting 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.

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 RSSM 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...
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:** GSSBA argumentation with the RS motion, respectively. Differently, a spline-based trajectory model to better reformulate the RS camera motion between consecutive frames is proposed by [19].

### 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 Normalization and covariance standardization Weighting 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 plug in existing GSSfM and GSSLAM solutions to handle the RS effect. The experiments show that NW-RSBA provides more accurate results and 10× 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 \( P_i = [X_i, Y_i, Z_i] \) is observed by a RS camera \( j \) represented by measurement \( m_{ij} \) in the image domain. The projection from 3D world to the image can be defined as:

\[
\begin{align*}
\mathbf{m}_{ij}^t &= [u_{ij}^t, v_{ij}^t]^T = \Pi([\mathbf{K}^t P_i^j]), \\
\mathbf{P}^j_i &= [X_i^j, Y_i^j, Z_i^j]^T = \mathbf{R}^j(v_i^j)P_i + \mathbf{t}^j(v_i^j),
\end{align*}
\]

where \( \Pi([X \ Y \ Z]^T) = \frac{1}{Z}[X \ Y]^T \) is the perspective division and \( \mathbf{K} \) is the camera intrinsic matrix [8]. \( \mathbf{R}^j(v_i^j) \in \text{SO}(3) \) 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:

\[
\begin{align*}
\mathbf{R}^j(v_i^j) &= (\mathbf{I} + [\omega^j]_x v_i^j)\mathbf{R}_0^j, \\
\mathbf{t}^j(v_i^j) &= \mathbf{t}_0^j + d^j v_i^j,
\end{align*}
\]
where $[\omega^j]^\top$ represents the skew-symmetric matrix of vector $\omega^j$ and $t_0^j$. $R_0^j$ is the translation and rotation matrix when the first row is observed. While $d^j = [d^j_x, d^j_y, d^j_z]^\top$ is the linear velocity vector and $\omega^j = [\omega^j_x, \omega^j_y, \omega^j_z]^\top$ is the angular velocity vector. Such model was adopted by [2,4,15]. Notice that RS instantaneous-motion is a function of $v^j_i$, 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 $K$ to the normalized measurement $[q_d^j, 1]^\top = K^{-1}[m_d^j, 1]^\top$. Thus, the projection model and camera instantaneous-motion become:

$$q_i^j = \begin{bmatrix} \rho_i^j \\ r_i^j \end{bmatrix}^\top = \Pi(R^j(r_i^j)\mathbf{P}_i + tv^j(r_i^j)), \quad (4)$$

$$R^j(r_i^j) = (I + [\omega^j]^\top r_i^j)R_0^j, \quad t^j(r_i^j) = t_0^j + d^j r_i^j. \quad (5)$$

In contrast to the direct-measurement-based RS model, $t_0^j$ and $R_0^j$ are the translation and rotation when the optical centre row is observed. $\omega^j$, $d^j$ 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 $\theta^*$ including camera poses $R^*$, $t^*$, instantaneous-motion $\omega^*$, $d^*$ and 3D points $\mathbf{P}^*$ by minimizing the reprojection error $e_i^j$ from point $i$ to camera $j$ over all the camera index in set $\mathcal{F}$ and corresponding 3D points index in subset $\mathcal{P}_j$:

$$\theta^* = (P^*, R^*, t^*, \omega^*, d^*) = \arg \min_\theta \sum_{j \in \mathcal{F}} \sum_{i \in \mathcal{P}_j} \|e_i^j\|^2. \quad (6)$$

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

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

**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 main differences 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.
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:

\[
\theta^* = \arg \min_\theta \sum_{j \in J} \sum_{i \in P_j} e_{ij}^T \Sigma_i^{-1} e_{ij},
\]

where \(e_{ij}\) is the reprojection error computed by normalized-measurement-based approach described in Eq. (4). By assuming the image measurement \(q_i^j\) follows a prior Gaussian noise: \(n_i^j \sim \mathcal{N}(0, W_i^j W_i^j^T)\), the newly introduced covariance matrix of the error \(\Sigma_i^j\) is defined as follows (proof in the supplemental material):

\[
\Sigma_i^j = C_i^j W_i^j W_i^j^T C_i^j^T, \tag{8}
\]

\[
C_i^j = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{z_i^j} \begin{bmatrix} 0 & -X_i^j \\ X_i^j & 0 \end{bmatrix} \begin{bmatrix} (\omega_i^j), R^j P_i + d_i^j \end{bmatrix}^T, \tag{9}
\]

\[
W_i^j = \begin{bmatrix} 1/f_x & 0 \\ 0 & 1/f_y \end{bmatrix}, \tag{10}
\]

where \(C_i^j\) and \(W\) 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 \(\Sigma_i^{-1}\), 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 3](image-url)  
(a) Prior image noise  
(b) NM-RSBA [2]  
(c) NW-RSBA

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

![Figure 5](image-url)  
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.
Algorithm 2: Solve the normal equation using two-stage Schur complement

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

**Output:** Update state vector $\delta$

1. Compute Schur complement matrix $S_p$ and $S_{rs}$ using Eq. (16) and (17);
2. Compute auxiliary vectors $t^*$ and $u^*$ using Eq. (19) and (18);
3. Solve Eq. (20) cascadingly:
   - Get $\delta_{rs}$ by solving $S_{rs}\delta_{rs} = -t^*$;
   - Get $\delta_{gs}$ by solving $U^*\delta_{gs} = -u^* - S^*\delta_{rs}$;
   - Get $\delta_p$ by solving $V\delta_p = -v - T^\top\delta_{rs} - W^\top\delta_{gs}$;
   - Stack $\delta_{gs}, \delta_{rs}, \delta_p$ into $\delta$;

\[ J^\top J = \begin{bmatrix} J_{rs}^\top J_{rs} & J_{rs}^\top J_{gs} & J_{rs}^\top J_p \\ J_{gs}^\top J_{gs} & J_{gs}^\top J_p \\ J_p^\top J_p \end{bmatrix} = \begin{bmatrix} R & S & T \\ S^\top & U & W \\ T^\top & W^\top & V \end{bmatrix} , \tag{14} \]

\[ J = \begin{bmatrix} J_{rs} & J_{gs} & J_p \end{bmatrix} = \begin{bmatrix} \frac{\partial \hat{e}}{\partial x_{rs}} & \frac{\partial \hat{e}}{\partial x_{gs}} & \frac{\partial \hat{e}}{\partial x_p} \end{bmatrix} , \tag{14a} \]

\[ \hat{e} = \{ \{ \hat{e}_r^j \} \} , \quad x_{gs} = \{ \{ R^j \} \} \{ t^j \} , \quad x_{rs} = \{ \{ \omega^j \} \} \{ d^j \} , \quad x_p = \{ \{ P^j \} \} , \tag{14b} \]

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

\[ \hat{e} = J^\top J \delta \]

\[ J^\top J \delta = \begin{bmatrix} \delta_{rs} \\ \delta_{gs} \\ \delta_p \end{bmatrix} = - \begin{bmatrix} t \\ u \\ v \end{bmatrix} = -J^\top \hat{e} , \tag{15} \]
where $\delta_{gs}$, $\delta_s$, and $\delta_p$ are the update state vectors to $x_{gs}$, $x_r$, and $x_p$, while $t$, $u$, and $v$ 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 $S_p$, $S_{rs}$ and two auxiliary vectors $u^*$ and $t^*$ to Eq. (15) as

$$
S_p = \begin{bmatrix}
R^* & S^* \\
S^* & U^*
\end{bmatrix},
S_{rs} = \begin{bmatrix}
R - TV^{-1}T^* & S - TV^{-1}W^* \\
S^* - WV^{-1}U^* & U - WV^{-1}W^*
\end{bmatrix},
\quad (16)
$$

$$
S_p = \begin{bmatrix}
R^* & S^* \\
S^* & U^*
\end{bmatrix},
S_{rs} = \begin{bmatrix}
R^* - S^*U^*-1S^*T \\
S^* - W^*V^{-1}U
\end{bmatrix},
\quad (17)
$$

$$
t^* = t - TV^{-1}v - S^*U^*-1u^*,
\quad (19)
$$

**Step 3: Orderly solve $\delta_{gs}$, $\delta_s$, and $\delta_p$.** Based on $S_p$, $S_{rs}$, $u^*$ and $t^*$, we reformulate the normal equation as

$$
\begin{bmatrix}
S_{rs} & 0 & 0 \\
S^* & U^* & 0 \\
T^* & W^* & V
\end{bmatrix}
\begin{bmatrix}
\delta_{gs} \\
\delta_s \\
\delta_p
\end{bmatrix}
= \begin{bmatrix}
t^* \\
u^* \\
v
\end{bmatrix},
\quad (20)
$$

which enables us to compute $\delta_s$ first, and then back substitutes the results to get $\delta_{gs}$. Finally, we can obtain $\delta_p$ based on the $3^{rd}$ 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 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).](image)

![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.](image)

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

<table>
<thead>
<tr>
<th>Seq</th>
<th>Duration [s]</th>
<th>length [m]</th>
<th>Realtime factor $\epsilon$</th>
<th>Tracking duration $DUR$</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>40</td>
<td>46</td>
<td>1.47</td>
<td>0.50</td>
</tr>
<tr>
<td>#2</td>
<td>27</td>
<td>37</td>
<td>1.51</td>
<td>0.90</td>
</tr>
<tr>
<td>#3</td>
<td>50</td>
<td>44</td>
<td>1.47</td>
<td>0.58</td>
</tr>
<tr>
<td>#4</td>
<td>38</td>
<td>30</td>
<td>1.61</td>
<td>1.1</td>
</tr>
<tr>
<td>#5</td>
<td>85</td>
<td>57</td>
<td>1.51</td>
<td>1</td>
</tr>
<tr>
<td>#6</td>
<td>43</td>
<td>51</td>
<td>1.47</td>
<td>0.76</td>
</tr>
<tr>
<td>#7</td>
<td>39</td>
<td>45</td>
<td>1.61</td>
<td>0.89</td>
</tr>
<tr>
<td>#8</td>
<td>53</td>
<td>46</td>
<td>1.56</td>
<td>0.79</td>
</tr>
<tr>
<td>#9</td>
<td>45</td>
<td>46</td>
<td>1.61</td>
<td>0.14</td>
</tr>
<tr>
<td>#10</td>
<td>54</td>
<td>51</td>
<td>1.56</td>
<td>0.78</td>
</tr>
<tr>
<td>t1-fast</td>
<td>28</td>
<td>50</td>
<td>1.92</td>
<td>1</td>
</tr>
<tr>
<td>t2-fast</td>
<td>29</td>
<td>53</td>
<td>1.92</td>
<td>1</td>
</tr>
</tbody>
</table>

4.2.1 RSSLM

We compare the performance of conventional GS-based $\text{Orb-SLAM}$ [18] versus augmented versions with $\text{NM-RSBA}$ [2] and proposed $\text{NW-RSBA}$ 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 $\text{Orb-SLAM+NW-RSBA}$ is slower than $\text{Orb-SLAM}$ by a factor roughly around 1.2 but is slightly faster than $\text{Orb-SLAM+NM-RSBA}$. As for tracking duration, $\text{Orb-SLAM}$ and $\text{Orb-SLAM+NM-RSBA}$ [2] fail with an average $DUR < 0.5$ in most sequences once the camera moves aggressively enough. In contrast, the proposed $\text{Orb-SLAM+NW-RSBA}$ achieves completed tracking with $DUR = 1$ in almost all sequences.

Absolute trajectory error. The ATE results on WHU-RSVI and TUM-RSVI datasets demonstrate that the proposed $\text{Orb-SLAM+NW-RSBA}$ is superior to $\text{Orb-SLAM}$ and $\text{Orb-SLAM+NM-RSBA}$ when dealing with RS effect. Qualitatively, this is clearly visible in Figs. 1 and 10. The sparse 3D reconstructions look much cleaner for $\text{Orb-SLAM+NW-RSBA}$ and close to the ground truth. The quantitative difference also becomes apparent in Tab. 3. $\text{Orb-SLAM+NW-RSBA}$ outperforms $\text{Orb-SLAM}$ and $\text{Orb-SLAM+NM-RSBA}$ 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 $\text{NW-RSBA}$ and compare quantitatively with related approaches. The results are presented in Tab. 4. The baseline methods’ performance show that $\text{NW-RSBA}$ obtains ATE of 0.007, which is half of the second best method $\text{DC-RSBA}$ and nearly 3% of $\text{GSBA}$. The removal of normalization from proposed $\text{NW-RSBA}$ adversely increases ATE up to 100%. The removal of covariance weighting from $\text{NW-RSBA}$ 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

1http://aric.whu.edu.cn/caolike/2019/11/05/the-whu-rsvi-dataset
2https://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, $DUR > 0.9$, $0.5 < DUR \leq 0.9$ and $DUR \leq 0.5$ are highlighted in light green, cyan, and orange.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GS data</td>
<td>Orb-SLAM [18]</td>
<td>#1 #2 #3 #4 #5 #6 #7 #8 #9 #10</td>
<td>t1-fast t2-fast</td>
</tr>
<tr>
<td>RS data</td>
<td>Orb-SLAM [18]</td>
<td>0.0015 0.013 0.018 0.107 0.030 0.013 0.054 0.053 0.020 0.024</td>
<td>0.044 0.008</td>
</tr>
<tr>
<td>RS data</td>
<td>Orb-SLAM+NM-RSBA [2]</td>
<td>0.059 0.100 0.411 0.126 0.055 0.044 0.237 0.218 0.176 0.373</td>
<td>0.237 0.018</td>
</tr>
<tr>
<td>RS data</td>
<td>Orb-SLAM+NW-RSBA (ours)</td>
<td>0.115 0.088 0.348 0.120 0.062 0.060 0.251 0.246 0.156 0.307</td>
<td>0.204 0.030</td>
</tr>
</tbody>
</table>

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 RSSiM 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.

<table>
<thead>
<tr>
<th>Ablation</th>
<th>Approach</th>
<th>ATE ($m$)</th>
<th>Runtime ($s$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSBA [16]</td>
<td>0.210</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>DC-RSBA [14]</td>
<td>0.016</td>
<td>1302</td>
<td></td>
</tr>
<tr>
<td>No normalization &amp; weighting</td>
<td>DM-RSBA [5]</td>
<td>0.023</td>
<td>740</td>
</tr>
<tr>
<td>No weighting</td>
<td>NM-RSBA [2]</td>
<td>0.020</td>
<td>15.8</td>
</tr>
<tr>
<td>No normalization</td>
<td>W-RSBA</td>
<td>0.013</td>
<td>16.1</td>
</tr>
<tr>
<td>No Schur complement with 1-stage Schur complement</td>
<td>NW-RSBA-0S</td>
<td>0.007</td>
<td>16.9</td>
</tr>
<tr>
<td>Consolidated</td>
<td>NW-RSBA-1S</td>
<td>0.007</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>NW-RSBA-2S</td>
<td>9.8</td>
<td></td>
</tr>
</tbody>
</table>

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


