SCAR: Satellite Imagery-Based Calibration for Aerial Recordings

Supplementary Material

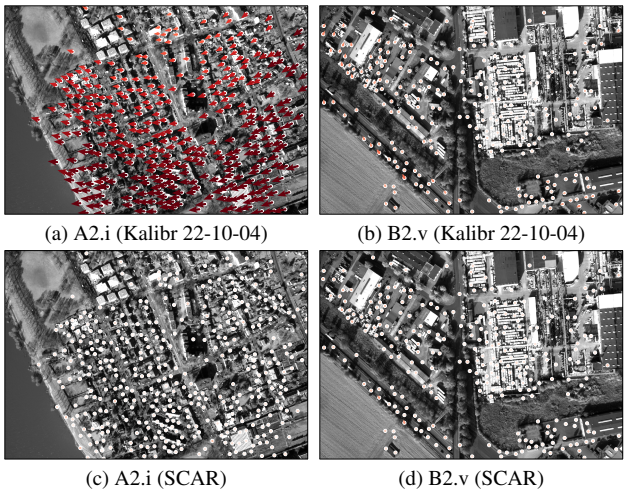
6. Qualitative Results

While the main paper presents a quantitative evaluation of SCAR, this section complements the analysis with qualitative examples that illustrate the effect of the calibration refinement in the image domain.

Two example images in Fig. 5 visualize the reprojection errors of automatically generated GCPs in the image domain. We select two complementary cases: sequence A2.i with large initial reprojection errors under Kalibr and sequence B2.v with comparatively small initial errors. In both cases, SCAR reduces the reprojection error.

For A2.i, the Kalibr calibration exhibits a pronounced roll-related misalignment, which is effectively corrected by SCAR. For B2.v, where the initial calibration already yields relatively small errors, SCAR still provides a consistent improvement. Besides systematic effects, residual errors in varying directions remain visible, reflecting noise originating from the image-based GCP generation. While these world-point uncertainties are partially reduced during optimization, they are not part of the calibration evaluation, as SCAR only estimates camera intrinsics and camera–INS extrinsics.

Overall, these qualitative examples illustrate the typical behavior observed across sequences and complement the quantitative evaluation presented in the main paper.



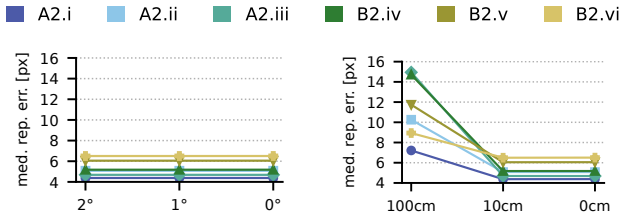
(a) A2.i (Kalibr 22-10-04) (b) B2.v (Kalibr 22-10-04)

(c) A2.i (SCAR) (d) B2.v (SCAR)

Figure 5. Qualitative reprojection comparison before and after SCAR refinement. The median reprojection error decreases from 47 px using the initial calibration (a) to 2 px after refinement (c) for A2.i, and, for B2.v, from 9 px (b) to 5 px (d).

Table 5. Overview of INS performance during training and validation sequences. Position accuracy is reported as mean positional stddev (x/y/z), orientation accuracy as mean angular stddev (r/p/y).

Seq. ID	GNSS type [%]	Position acc. [m]	Orientation acc. [°]
A1.i	0 / 100	0.04 / 0.03 / 0.05	0.07 / 0.08 / 0.16
A1.ii	0 / 100	0.05 / 0.04 / 0.07	0.07 / 0.07 / 0.14
A1.iii	0 / 100	0.06 / 0.04 / 0.05	0.08 / 0.08 / 0.14
B1.iv	0 / 100	0.02 / 0.03 / 0.07	0.08 / 0.08 / 0.16
B1.v	0 / 100	0.07 / 0.03 / 0.16	0.08 / 0.08 / 0.16
B1.vi	12 / 88	0.08 / 0.09 / 0.11	0.08 / 0.08 / 0.20
A2.i	0 / 100	0.04 / 0.03 / 0.06	0.07 / 0.07 / 0.14
A2.ii	0 / 100	0.04 / 0.04 / 0.08	0.07 / 0.07 / 0.14
A2.iii	0 / 100	0.04 / 0.06 / 0.07	0.08 / 0.08 / 0.15
B2.iv	0 / 100	0.04 / 0.04 / 0.14	0.07 / 0.07 / 0.15
B2.v	0 / 100	0.07 / 0.03 / 0.11	0.08 / 0.07 / 0.14
B2.vi	0 / 100	0.09 / 0.07 / 0.11	0.07 / 0.07 / 0.14



(a) White rotation pose noise (b) White translation pose noise

Figure 6. Analysis on the effect of imprecise reference poses.

7. Pose Robustness Evaluation

In the main paper, we assume accurate INS poses during calibration in GNSS-enabled sequences. To provide a clearer picture of the actual pose quality used in our experiments, this section summarizes the accuracy of the INS measurements across all training and validation sequences.

Table 5 reports the GNSS solution types together with the corresponding positional and angular accuracy estimates provided by the onboard INS. The GNSS solution type is given as the ratio of non-RTK to RTK measurements within each sequence. No GNSS outages occurred during the sequences.

The reported position uncertainties correspond to the mean positional standard deviations in the x, y, and z directions as estimated by the onboard extended Kalman filter (EKF) of the INS. These estimates were additionally verified using post-processed kinematic (PPK) GNSS processing. Orientation uncertainties correspond to the mean EKF standard deviations of roll, pitch, and yaw.

As can be observed, the INS provides centimeter-level

positional accuracy in nearly all sequences due to continuous RTK GNSS availability. The employed dual-antenna setup, together with the integrated IMU and magnetometer whose measurements are fused within the EKF, further provides highly accurate orientation estimates. These precise reference poses enable reliable calibration and validation and allow SCAR to produce accurate calibration parameters that can later be applied in GNSS-denied scenarios.

To investigate the sensitivity of SCAR to degraded reference poses, we additionally perform a robustness analysis in which artificial noise is added to the INS poses before calibration. Specifically, we perturb the orientation of each pose by independent Gaussian noise with standard deviations of 1° and 2° applied to each axis. Similarly, translation noise with standard deviations of 10 cm and 100 cm is added independently to the position components. The resulting calibration performance is shown in Figures 6a and 6b.

The results indicate that orientation noise has only a minor influence on the final calibration outcome. This can be explained by the fact that rotational components are well constrained in the visual localization setting, allowing the camera orientation to be accurately estimated even when the INS poses are locally perturbed. Since the injected noise is independent and does not introduce a systematic bias, the errors tend to average out during optimization.

In contrast, translation noise leads to a noticeable degradation in calibration performance. As also observed in the visual localization experiments in the main paper, translation is significantly harder to constrain than rotation. Consequently, position errors in the reference poses cannot be fully compensated by the optimization, resulting in biased calibration estimates when large translation errors are present.

8. Camera Calibration Optimization

The primary residual terms of the factor graph are introduced in the main paper. For completeness, we briefly define the camera calibration residual here, although it follows the standard formulation used in bundle adjustment.

Camera intrinsics are modeled using the standard pinhole projection with radial–tangential distortion with parameters

$$K = (f_x, f_y, c_x, c_y, k_1, k_2, p_1, p_2)$$

consisting of focal lengths, principal point, radial distortion coefficients, and tangential distortion coefficients.

Given a 3D world point X_j observed from the camera pose $C_t = (R_t, p_t)$, the point is first transformed into the camera coordinate frame by $x_{t,j}^{\text{cam}} = R_t^\top (X_j - p_t)$. Normalized image coordinates are obtained as $x = {}^{x_{t,j}^{\text{cam}}}/_{z_{t,j}^{\text{cam}}}$ and $y = {}^{y_{t,j}^{\text{cam}}}/_{z_{t,j}^{\text{cam}}}$, and radial–tangential distortion is applied as $r^2 = x^2 + y^2$. Finally, distorted normalized coordinates are mapped to pixel coordinates

$$\pi(K, C_t, X_j) = \begin{pmatrix} f_x x_d + c_x \\ f_y y_d + c_y \end{pmatrix},$$

with x_d, x_d following Brown–Conrady model. The reprojection residual for observation $u_{t,j}$ is therefore

$$r_{t,j}^{\text{rep}} = u_{t,j} - \pi(K, C_t, X_j).$$

These residuals are incorporated into the BA factors of the optimization. Analytic Jacobians regarding the intrinsic parameters K, pose C_t, and landmark position X_j are provided by the underlying GTSAM implementation.