Additional evaluations for VI-HSO: Hybrid Sparse Monocular Visual-Inertial Odometry

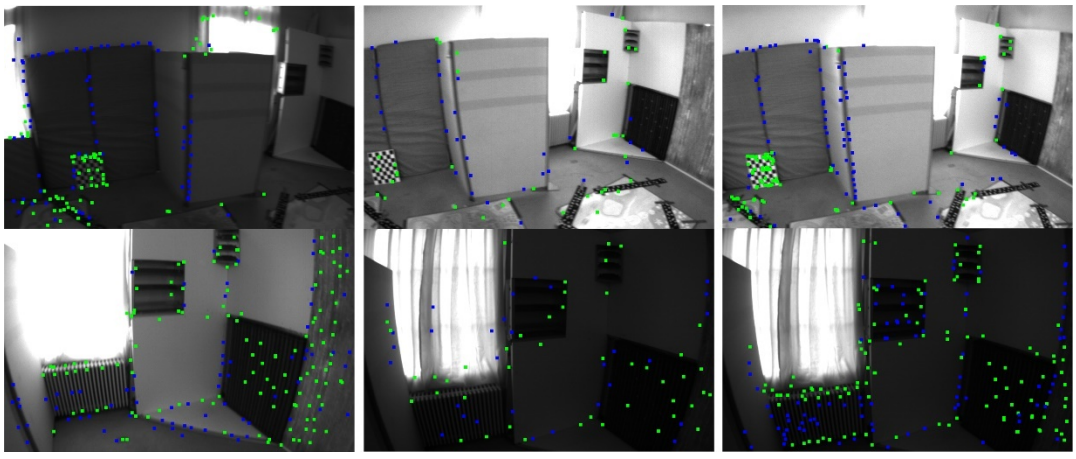
Wenzhe Yang, Yan Zhuang, Dongting Luo, Wei Wang, and Hong Zhang

We provide test results of AIA and DIDF under different situations in section I, comparative experiments on the two rules of AIA module in section II, comparative experiments on the two dynamic ranges of DIDF module in section III, a detailed ablation study on AIA module in section IV, and comparative experiments between VI-HSO and HSO in section V.

I. Testing AIA and DIDF under different situations

We conducted a detailed analysis of the impact of the AIA module on the system under conditions of significant intensity changes and rapid rotation. Additionally, we conducted comparative experiments on the DIDF module in motion blur and texture-less regions. The number of tracked feature points is used as an evaluation criterion in the comparative experiments. A large number of tracked feature points can be expected to lead to accurate pose estimation and high system robustness.

(1) Testing AIA under significant changes in image intensity

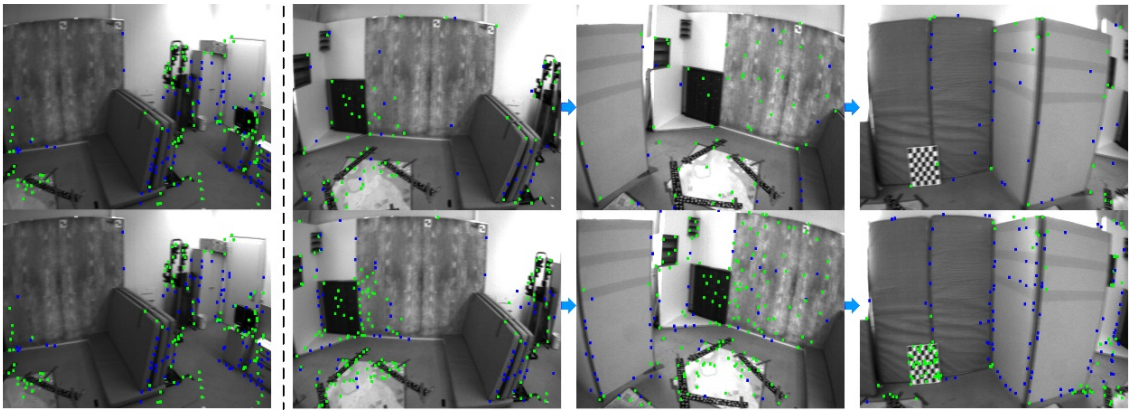


(a) Reference frame (b) Current frame without AIA (c) Current frame with AIA

Fig. 1. Feature tracking under two opposite image intensity changes. In the first row, the image becomes brighter. In the opposite case of the second row, the image becomes darker. (a) is the reference frame. (b) is the feature tracking in the current frame without AIA. (c) is the result of AIA, and the green and blue points are the feature points tracked.

We conducted comparative experiments on two situations of image brightening and darkening. As shown in Fig. 1, the number of feature points tracked in the current frame significantly decreases without the AIA module, when the image brightness changes dramatically. The number of feature points tracked by our method is well maintained.

(2) Testing AIA under rapid rotation

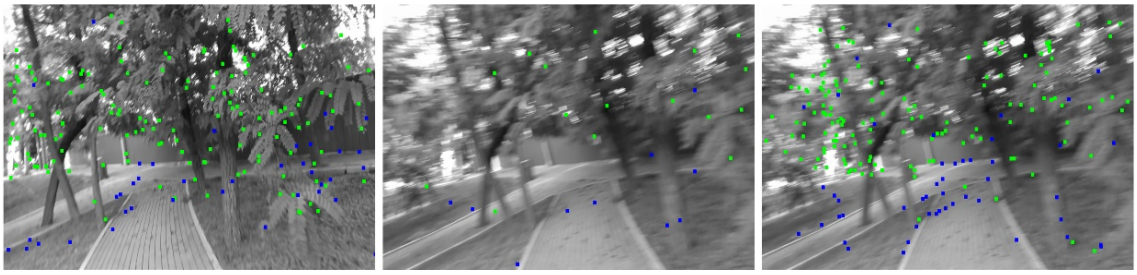


(a) Reference frame (b) Continuous rapid rotation frames

Fig. 2. Feature tracking under rapid rotation. The first row is the rotation process without AIA, while the second row is with AIA. (a) is the reference frame. (b) represents the continuous rapid rotation frames. The green and blue points are the feature points tracked.

As shown in Fig. 2, we carried out comparative tests on feature tracking by the system under rapid rotation. As the camera rotates rapidly, the number of feature points tracked gradually decreases without the AIA module. The system with the AIA module could track more feature points. This indicates that our system can well estimate the interframe motion under rapid rotation, allowing more feature points to be projected to subsequent frames.

(3) Testing DIDF under motion blur

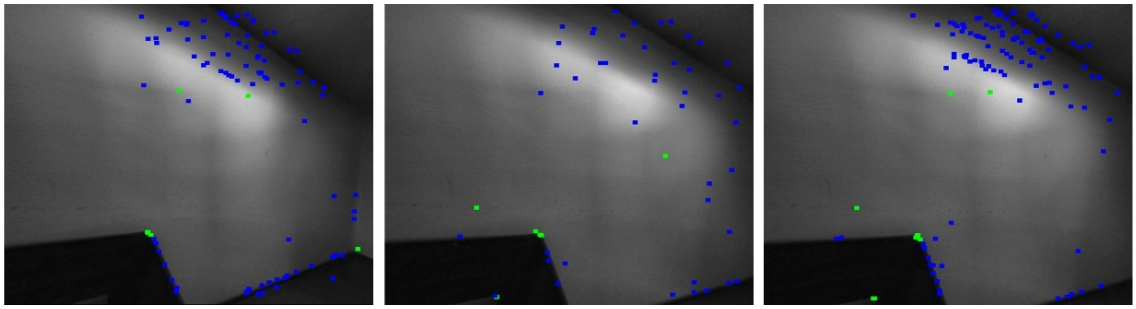


(a) Reference frame (b) Current frame without DIDF (c) Current frame with DIDF

Fig. 3. Feature tracking under motion blur. (a) is the reference frame. (b) is the feature tracking in the current frame without DIDF. (c) is the result of DIDF, and the green and blue points are the feature points tracked.

The comparative experiments with and without the DIDF module were made in motion blur. As shown in Fig. 3, only a few feature points can be tracked without the DIDF module. Our method can track more feature points, improving the robustness of the system with respect to motion blur.

(4) Testing DIDF in texture-less regions



(a) Reference frame (b) Current frame without DIDF (c) Current frame with DIDF

Fig. 4. Feature tracking in texture-less regions. (a) is the reference frame. (b) shows the feature tracking in the current frame without DIDF. (c) shows the result of DIDF. The green and blue points are the feature points tracked.

As shown in Fig. 4, we conducted comparative experiments on the DIDF module in texture-less regions. It is obvious that more feature points can be tracked with the DIDF module, which is beneficial for improving the robustness of the system.

II. COMPARATIVE EXPERIMENTS ON THE TWO RULES OF AIA

We conducted comparative experiments on the two rules of AIA module proposed in Eq. (6). The MH05 sequence on EuRoC dataset was used as the test scenario because there were areas with significant brightness changes and rapid rotational motion in the scene.

TABLE I compares our rule to two other separate rules and our rule outperforms others. As shown in Fig. 5, the pixel gradient rule can effectively improve system accuracy when there is a significant change in image intensity. The rotation rule can better ensure the accuracy of interframe motion estimation.

TABLE I

TRANSLATION ERROR [M] ON MH05 SEQUENCE

	Pixel Gradient Only	Rotation Only	Ours
MH05	0.093	0.122	0.067

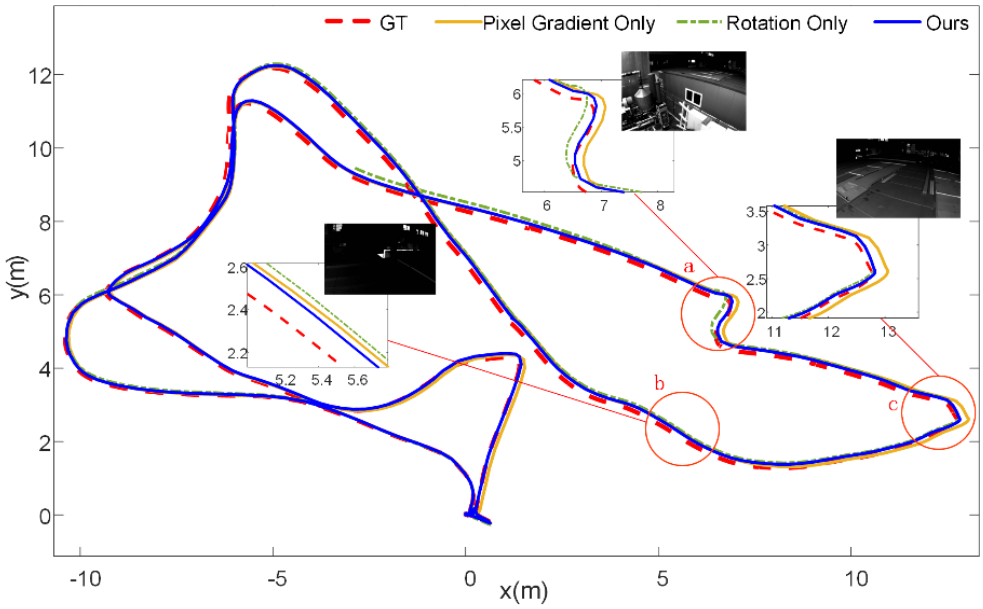


Fig. 5. Trajectories comparison results of the MH05 sequence on EuRoC dataset. Red circle (a) is the area where the camera rotates rapidly and the brightness of the image changes significantly. Our rule is significantly better than the other two rules. Red circle (b) is the area with poor lighting. The trajectory error of the pixel gradient rule is smaller than the error of the rotation rule. Red circle (c) is an area of the rapid camera rotation. The trajectory of the rotation rule is closer to the ground truth (GT) than the trajectory of the gradient rule.

III. COMPARATIVE EXPERIMENTS ON THE TWO DYNAMIC RANGES OF DIDF

We conducted comparative experiments on the two dynamic ranges (L1 and L2) proposed in DIDF. For the candidate points extracted from the current frame, L1 describes continuous frame constraints, and L2 represents covisibility frame constraints. The V203 sequence on EuRoC dataset and the Grove sequence on the self-collected real-word dataset were used as the test scenarios. In TABLE II, we compare our method under the two constraints respectively and our method clearly performs the best.

TABLE II
TRANSLATION ERROR [M] ON V203 AND GROVE SEQUENCE

	L1 Only	L2 Only	Ours
V203	0.047	0.052	0.029
Grove	0.656	0.347	0.220

On V203, L1 only performs better than L2 only. This is because in texture-less regions and under fast motion, due to the presence of the IMU, continuous frame estimation is relatively accurate and can better constrain the convergence of candidate points, as shown in Fig.6.

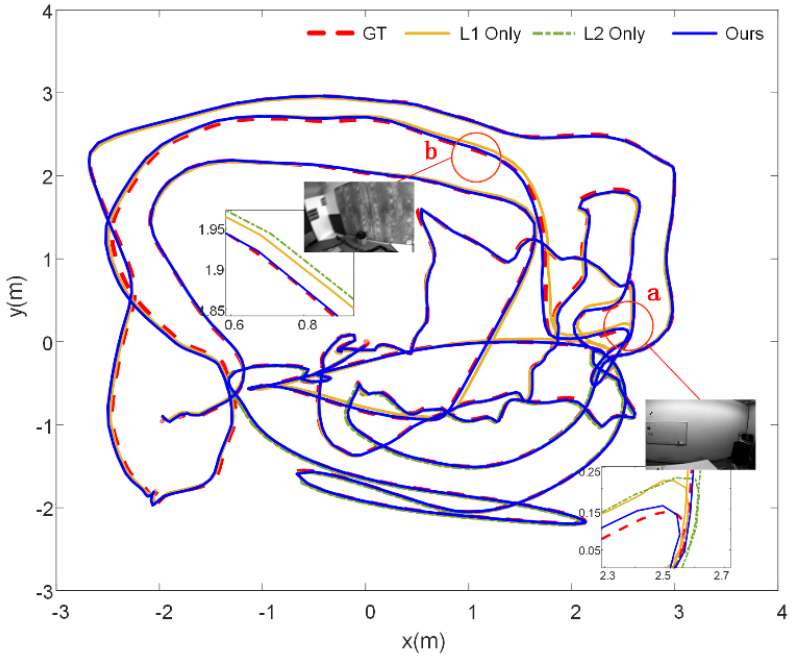


Fig. 6. Trajectories comparison results of the V203 sequence on EuRoC dataset. Red circle (a) corresponds to a texture-less region. Red circle (b) represents camera frames involving fast motion. Our method performs the best, and L1 only is closer to the ground truth than L2 only.

The effect of L2 only is better than that of L1 only on Grove. In the case of motion blur and partial occlusion, covisibility frames can provide more constraints for candidate points, making it easier to converge, as shown in Fig. 7.

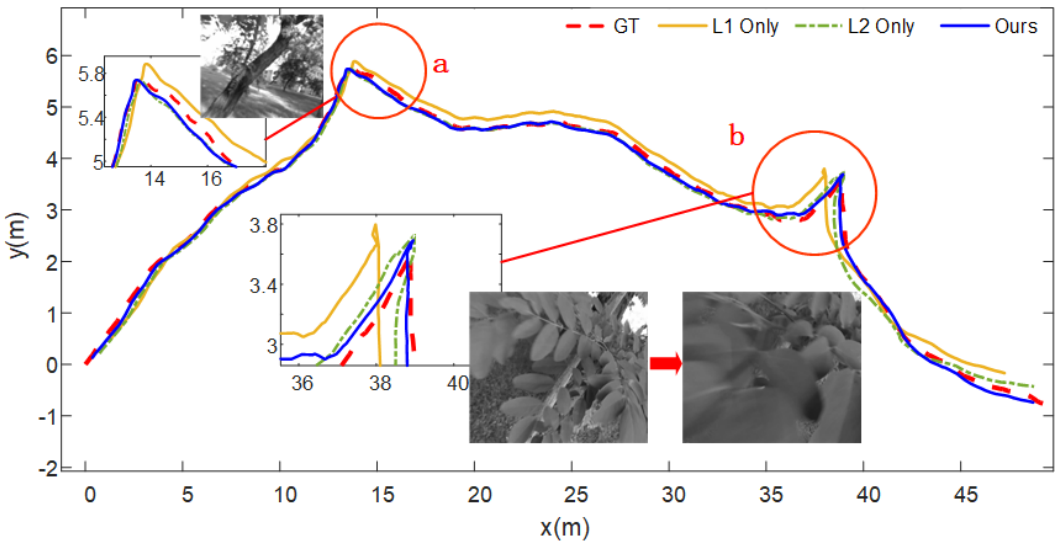


Fig. 7. Trajectory comparison results of the Grove sequence on self-collected real-word dataset. Red circle (a) is the motion blur area. Red circle (b) represents the situation where the image is partially occluded. Our method is the closest to the ground truth. Compared to L2 only, L1 only has greater drift.

IV. Ablation AIA Module

In Table III, we conducted experiments on the EuRoC dataset to compare the accuracy of the conventional image alignment, inverse compositional method, method of IMU estimation, and our method (dynamic switching method). And we tested the running time of each module on the dataset.

TABLE III
TRANSLATION ERROR [M] AND CORRESPONDING MODULE
MEAN RUNNING TIME [MS]OF EACH SEQUENCE ON EUROC DATASET

	MH01	MH02	MH03	MH04	MH05	V101
Conventional image alignment	0.035	0.034	0.061	0.066	0.067	0.036
Inverse compositional	0.033	0.036	0.053	0.072	0.082	0.036
IMU	0.043	0.041	0.068	0.074	0.071	0.036
Ours	0.021	0.033	0.041	0.047	0.062	0.035

	V102	V103	V201	V202	V203	Running time
Conventional image alignment	0.024	0.028	0.032	0.021	0.047	4.58±1.15
Inverse compositional	0.020	0.033	0.026	0.029	0.052	2.14±0.65
IMU	0.022	0.026	0.029	0.024	0.062	0.002±0.001
Ours	0.019	0.023	0.025	0.020	0.023	3.32±0.76

V. Comparative Experiments Between VI-HSO and HSO

We conducted comparative experiments between VI-HSO and HSO on the EuRoC dataset, as shown in Table IV. Our method has significantly better accuracy than HSO on the same dataset.

TABLE IV
TRANSLATION ERROR [M] OF EACH SEQUENCE ON EUROC DATASET

Sequence	MH01	MH02	MH03	MH04	MH05	V101
HSO	0.063	0.095	0.102	0.190	0.152	0.077
VI-HSO (Ours)	0.021	0.033	0.041	0.047	0.062	0.035

Sequence	V102	V103	V201	V202	V203	**Avg**
HSO	0.046	0.273	0.056	0.114	0.354	0.139
VI-HSO (Ours)	0.019	0.023	0.025	0.020	0.023	0.032