SDGS: Spatial Difference Guided Gaussian Splatting for Simultaneous Localization and 3D Reconstruction

Yijian Tian Mingtao Ou Zijian Pan Xinglong Ji*
Center for Brain-Inspired Computing Research (CBICR),
Department of Precision Instrument, Tsinghua University

Email: yijian00@outlook.com (Yijian Tian)
*Corresponding author. Email: xinglongji@mail.tsinghua.edu.cn

SDGS: Spatial Difference Guided Gaussian Splatting for Simultaneous Localization and 3D Reconstruction

Supplementary Material

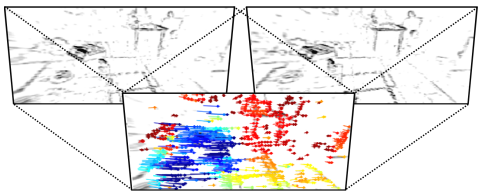


Figure 1. Sparse stereo disparity estimation by solving an LK-style problem. Top: stereo SD pairs (polarity omitted). Bottom: disparity results, where blue indicates larger disparity and red indicates smaller disparity.

1. Supplementary Material

1.1. Hybrid Pixel Vision Sensor

How machines can accurately perceive the world has been a long-standing topic in computer vision. From traditional RGB cameras with Bayer filters that capture color, to event cameras that offer extremely high frame rates and dynamic range, many different sensor types have been explored. However, RGB cameras require a finite exposure time, which makes them sensitive to high-speed motion and HDR scenes; event cameras, in contrast, can easily capture fast motion but are unable to sense static scenes.

A brain-inspired hybrid pixel vision chip [4] combines the advantages of RGB and event cameras in a chessboard-like layout. It provides two pathways for multimodal sensing: a cognition-oriented pathway (RGB) for scene understanding, and an action-oriented pathway for fast motion perception, which is similar to traditional event cameras while additionally providing spatial-difference (SD) perception. SD computes the voltage difference between neighboring pixels in the action-oriented pathway, yielding spatial gradients at the frame rate of an event camera but without requiring motion, unlike traditional event cameras. This design is analogous to the rods and cones in the human eye, which respectively perceive low-light motion and color, forming complementary pathways. The complementary design of SD and RGB is naturally compatible with our proposed algorithm, enabling high-fidelity 3D reconstruction and camera localization under extremely fast motion in real-world scenarios.

1.2. LK Disparity

A simple demonstration of our sparse disparity method based on [3] is shown in Fig. 1.

1.3. Camera Trajectories on the 2D Plane

To better understand the motion patterns of the evaluated sequences, we visualize the camera trajectories on the 2D plane using evo [2] as shown in Fig. 2 and Fig. 3. For the stereo-Tianmouc dataset, we select three sequences with gradually increasing motion difficulty: tianmouc/slow, tianmouc/fast, and tianmouc/extreme. For the TUM RGB-D dataset, we use three representative sequences: fr1/desk, fr2/xyz, and fr3/office. These top-down trajectory plots provide an intuitive picture of the motion scale and trajectory complexity of each sequence.

1.4. Datasets Quantification

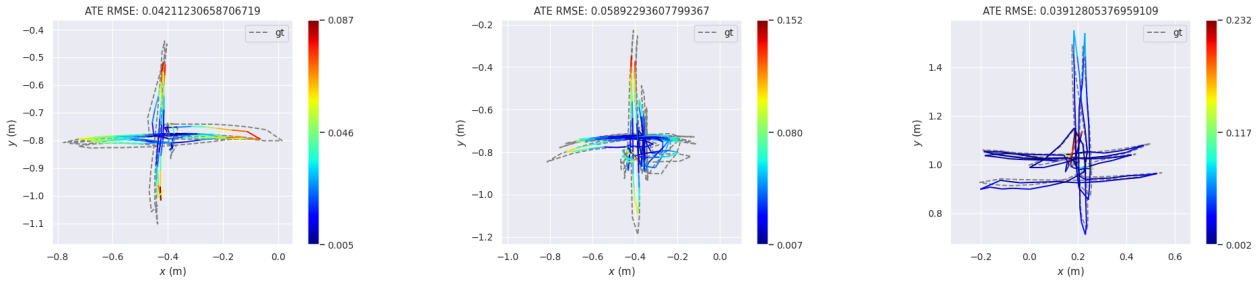
1.4.1. Stereo-Tianmouc

Ground truth (GT) is derived from camera-mounted markers via a motion capture system; hand-eye calibration computes the transformation matrix between markers and cameras' optical center. The camera is calibrated using its RGB stream by traditional methods, and SD intrinsics are derived from RGB intrinsics (differing only in resolution). SD and RGB streams (both hardware-level outputs) are hardware-timestamp synchronized directly.

To quantify the amount of motion in the stereo-Tianmouc dataset, we compute dense Farnebäck optical flow [1] between consecutive RGB frames. For each frame pair, we record the mean and maximum flow magnitudes (in pixels), and report their distributions for three subsets: tianmouc/slow, tianmouc/fast, and tianmouc/extreme. On tianmouc/slow, the average mean motion is 2.34 px/frame, with a global maximum displacement of 47.72 px/frame. In contrast, tianmouc/fast and tianmouc/extreme exhibit significantly larger motion: tianmouc/fast has an average mean motion of 7.51 px/frame and a maximum displacement of 119.19 px/frame, while tianmouc/extreme reaches 6.70 px/frame on average with a maximum displacement of 193.28 px/frame. The corresponding distributions of mean and maximum motion are summarized in Fig. 4. These statistics confirm that the tianmouc/fast and tianmouc/extreme sequences contain large inter-frame motion and strong motion blur, where SD information becomes particularly valuable. A brief visualization of each RGB sequence is shown in Fig. 5.

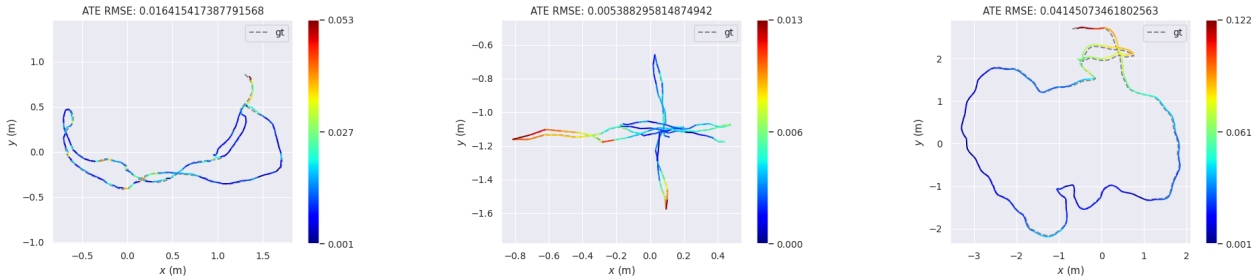
1.4.2. SD-Replica

For the SD-Replica dataset, we generate SDL, which corresponds to SD in the upper-left and lower-left directions, and SDR, which corresponds to SD in the upper-right and



(a) tianmouc/slow (b) tianmouc/fast (c) tianmouc/extreme

Figure 2. Top-down camera trajectories on the stereo-Tianmouc dataset.



(a) fr1/desk (b) fr2/xyz (c) fr3/office

Figure 3. Top-down camera trajectories on the TUM RGB-D dataset. The three sequences cover small desktop motion (fr1/desk), larger translational motion (fr2/xyz), and cluttered office motion (fr3/office).

lower-right directions. Both are polarized and quantized by an 8-bit ADC. Motion blur is synthesized by integrating the appearance of the static 3D scene along the ground-truth camera trajectory over a finite exposure time. A visualization is shown in Fig. 6.

1.5. Limitations

Although our method demonstrates efficient use of sparse edges and robust reconstruction with the hybrid pixel sensor, there is still room for improvement. For example, a sparse 3DGS optimizer could substantially accelerate both tracking and mapping, and thus further improve tracking accuracy. Currently, we are using linearized approximation for pinhole model Gaussian projection, which introduces linearization error. In addition, algorithms that explicitly model inconsistent edges or noise, which frequently appear in SD images, would likely improve the overall robustness of the system.

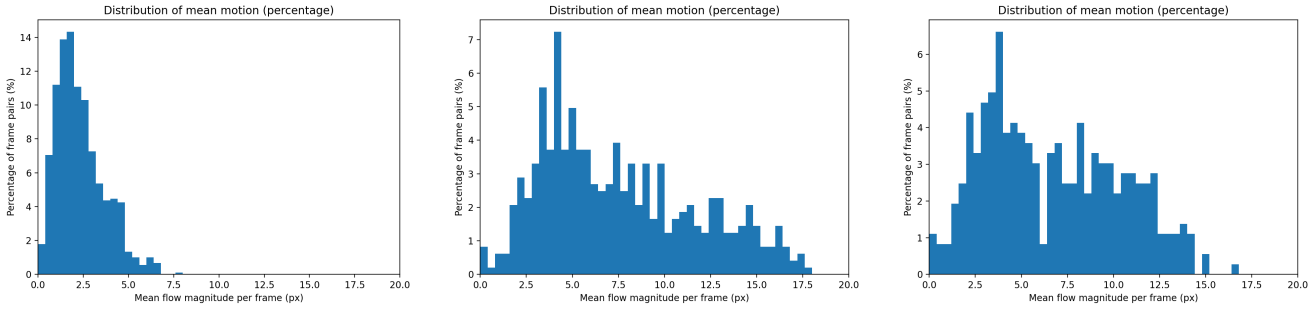
References

[1] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis*, pages 363–370. Springer, 2003. 1

[2] Michael Grupp. evo: Python package for the evaluation of odometry and slam. https://github.com/MichaelGrupp/evo, 2017. 1

[3] Bruce D Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI'81: 7th international joint conference on Artificial intelligence*, pages 674–679, 1981. 1

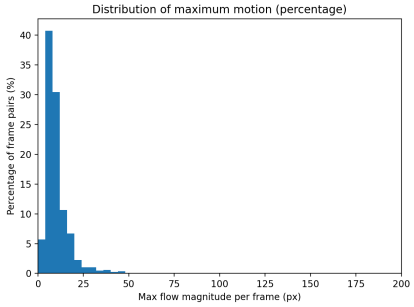
[4] Zheyu Yang, Taoyi Wang, Yihan Lin, Yuguo Chen, Hui Zeng, Jing Pei, Jiazheng Wang, Xue Liu, Yichun Zhou, Jianqiang Zhang, et al. A vision chip with complementary pathways for open-world sensing. *Nature*, 629(8014):1027–1033, 2024. 1



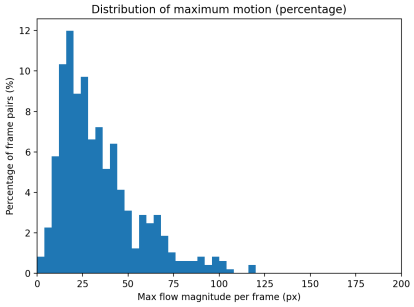
(a) Slow subset (mean motion)

(b) Fast subset (mean motion)

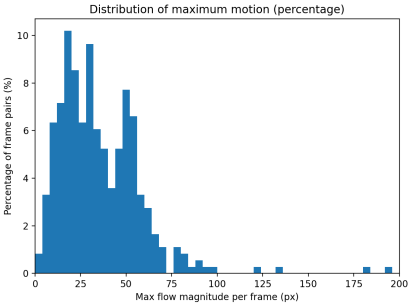
(c) Extreme subset (mean motion)



(d) Slow subset (max motion)



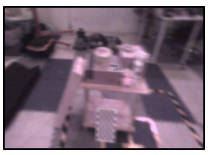
(e) Fast subset (max motion)



(f) Extreme subset (max motion)

Figure 4. Distributions of inter-frame motion for the stereo-Tianmouc datasets. Top row: histograms of the mean flow magnitude per frame pair. Bottom row: histograms of the maximum flow magnitude per frame pair.

(a) tianmouc/slow

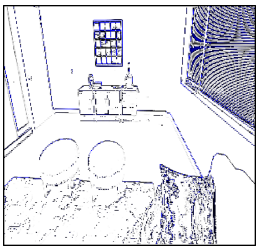


(b) tianmouc/fast

(c) tianmouc/extreme

Figure 5. Stereo-Tianmouc dataset RGB sequences under different motion speeds.

(a) Blurred RGB



(b) SDL

(c) SDR

Figure 6. Examples from the SD-Replica dataset.