

---

# Supplementary Material for Event-based HDR Structured Light

---

Jiacheng Fu Yue Li Xin Dong Wenming Weng Yueyi Zhang Zhiwei Xiong\*  
University of Science and Technology of China  
jc\_fu@mail.ustc.edu.cn zwxiong@ustc.edu.cn

This supplementary material provides additional details and results to support our main paper. Section 1 introduces the 3D reconstruction pipeline of the event-based structured light (SL) system. Section 2 presents samples from our proposed synthetic and real-world benchmarking datasets. Section 3 showcases additional qualitative results, which convincingly demonstrate the robustness of our method in reconstructing challenging surfaces under diverse conditions. Section 4 further demonstrates the universality of our confidence-driven stereo matching strategy, which achieves consistent improvements when applied to two representative stereo models on both synthetic and real-world datasets.

## 1 3D Reconstruction from Event-based Structured Light System

We adopt a traditional monocular structured light (SL) setup based on random speckle coding. As shown in Fig. 1(a), our system is comprised of a Prophesee event camera [1] and a digital light projector (DLP) [2]. In operation, the projector sequentially projects our designed multi-frame HDR coding patterns to the scene, and the illuminance changes are captured by the event camera and generate the corresponding event stream, as shown in Fig. 1(b). Further, we adopt a similar method used by [3] for event stream division and event framing, which generates event frames with scene depth encoding.

For disparity estimation, we model the projector as an inverse camera and treat the projected patterns as its captured images, which we call the reference frames. We then perform epipolar rectification between the captured event frames and the reference frames, allowing the use of generic stereo matching algorithms for disparity estimation. The disparity can easily be converted to depth based on the calibration parameters. This pipeline facilitates reliable 3D reconstruction.

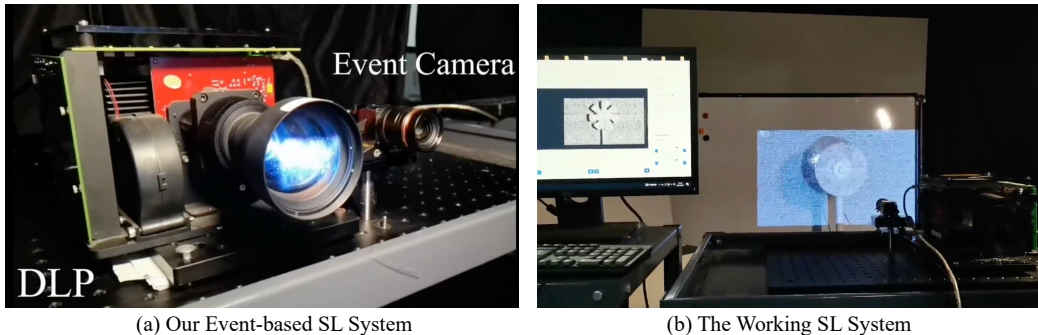


Figure 1: Our built event-based SL system.

---

\*Corresponding author

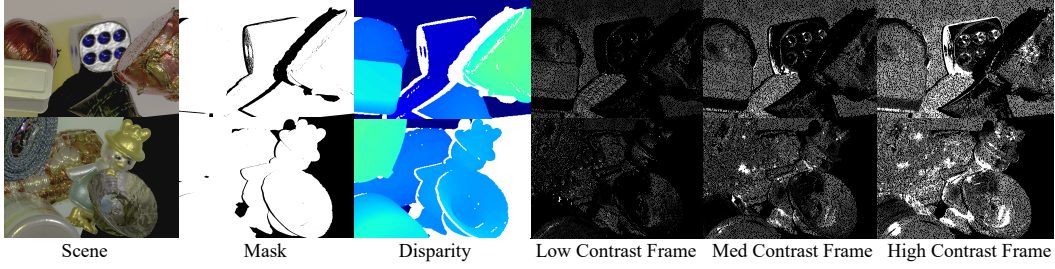


Figure 2: Samples from the synthetic dataset.

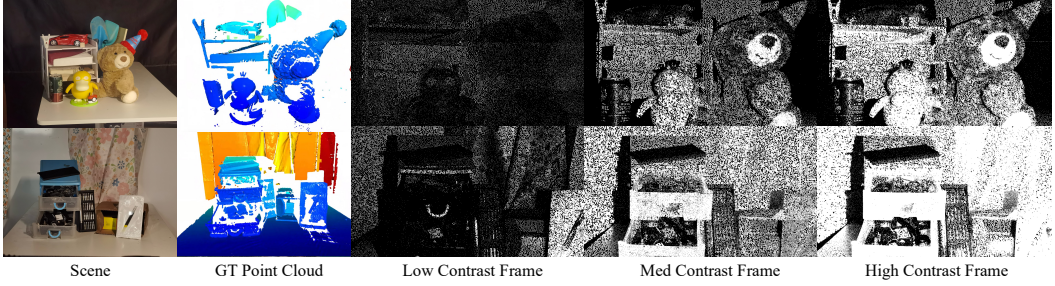


Figure 3: Samples from the real-world dataset.

## 2 Dataset Samples

**Synthetic Dataset.** Fig. 2 shows rendered samples from our synthetic dataset, each sample contains an RGB frame of the scene, a mask indicating regions that are covered by the random speckle, ground truth (GT) disparity and low-, medium-, high-contrast event frame. It is worth noting that, the mask rendered is used in supervised training to avoid the supervision of regions not covered by random speckles, which draws ambiguity in training. As can be seen, the constructed scenes have significant diversity in object categories, textures, surface properties, and scene lighting conditions, which allows the shape and distribution of rendered speckles to closely approximate those in real-world scenarios. As a result, the model trained on this high-quality synthetic dataset generalizes well to real-world scenes.

**Real-World Dataset.** Fig. 3 shows collected samples from our real-world dataset. Each sample contains three event frames generated by our multi-contrast coding and GT point cloud captured by Photoneo MotionCam-3D M+ [4].

## 3 Additional Synthetic and Real-World Results

**Synthetic Results.** More qualitative results on the synthetic dataset are shown in Fig. 4. In the figure, the white regions indicate areas not covered by the projected speckle patterns, which are manually excluded using the rendered mask. In sample 1, low reflectivity causes significant speckle absence, nevertheless, our method demonstrates a strong capability of disparity inference and completion, effectively minimizing mismatched areas. In sample 2, our method also demonstrates the ability to reconstruct highly reflective regions. Besides, our method produces smoother disparity estimation while retaining the ability to reconstruct fine details, as evidenced by the reconstruction of the pizza slice surface.

**Real-World Results.** More qualitative results on the real-world dataset are shown in Fig. 5. Among all compared methods, only ours can produce accurate and sharp disparity estimates in complex HDR scenes. Moreover, our method demonstrates the ability to reconstruct transparent objects in both samples, highlighting the robustness of our method in dealing with non-Lambertian surfaces.

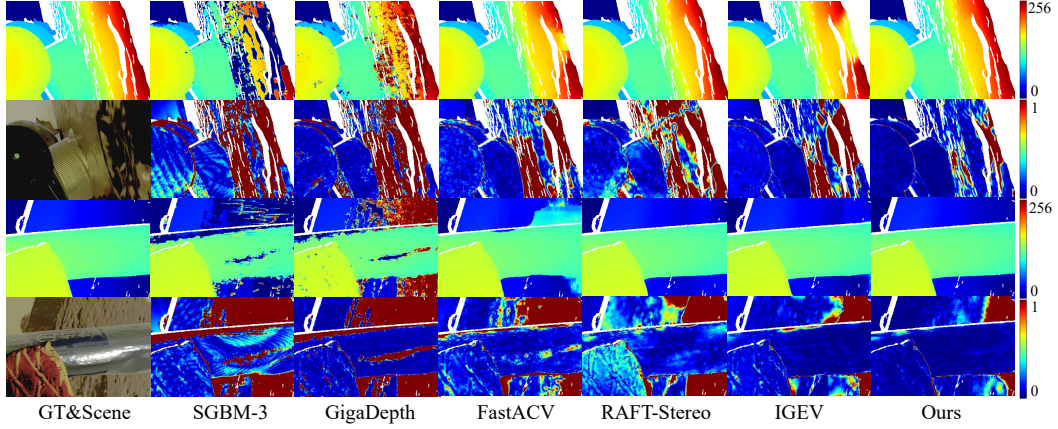


Figure 4: Additional qualitative results on the synthetic dataset. Two samples are shown, with the first row showing the disparities and the second row showing the scene and the error maps.

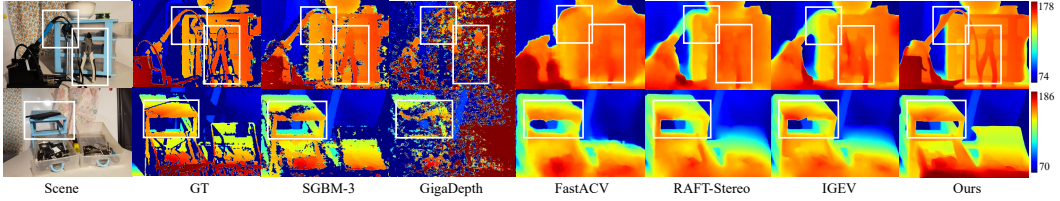


Figure 5: Additional qualitative results on the real-world dataset.

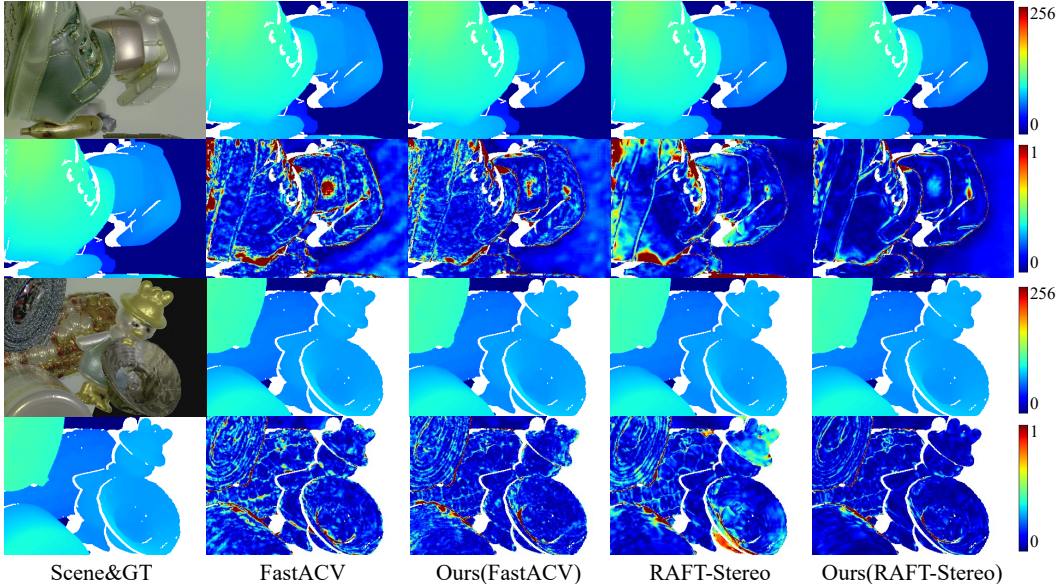


Figure 6: Qualitative evaluation of the universality of our confidence-driven stereo matching strategy on the synthetic dataset.

#### 4 Universality of Confidence-Driven Stereo Matching Strategy

To further validate the universality of the proposed confidence-driven stereo matching strategy across different models, we conduct additional qualitative experiments on both synthetic and real-world datasets. We apply our proposed strategy to two representative stereo matching algorithms: FastACV [5] and RAFT-Stereo [6]. The results on the synthetic dataset are presented in Fig. 6. As can be seen, by addressing inter-frame interference, our strategy significantly reduces mismatches while enhancing local matching accuracy. The results on the real-world dataset are presented in Fig.



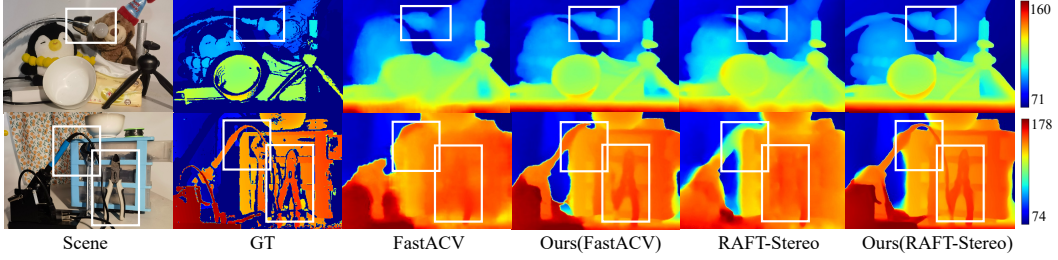


Figure 7: Qualitative evaluation of the universality of our confidence-driven stereo matching strategy on the real-world dataset.

7. Both baseline models produce blurred reconstructions due to their inability to identify regions with valid coding across the three frames. With our proposed confidence-driven strategy, both two baseline models can reconstruct HDR scenes with significantly improved quality.

## 5 Comparison with a General-Purpose Event-based SL Method

To demonstrate the advantage of our method in HDR scenes, we implement a state-of-the-art event-based SL method that is based on graycode [7], and make comparison with the proposed method in two real-world HDR scenes. Quantitative results are shown in Table. 1. As can be seen, [7], which is designed for general-scene reconstruction, fails in HDR scenes. Our proposed HDR 3D reconstruction framework (only 3 frames) significantly outperforms [7] (requiring 11 frames) in terms of both accuracy and completeness.

Scene	Method	EPE ↓	Bad 0.5 ↓	Bad 1.0 ↓	Bad 2.0 ↓	Bad 3.0 ↓	Bad 5.0 ↓	D1 ↓
Scene 1	Graycode	43.6115	0.7982	0.6407	0.4766	0.4424	0.4315	0.4182
	Ours	<b>0.8449</b>	<b>0.2727</b>	<b>0.0781</b>	<b>0.0417</b>	<b>0.0379</b>	<b>0.0286</b>	<b>0.0249</b>
Scene 2	Graycode	38.7694	0.7380	0.5921	0.4387	0.4068	0.3952	0.3810
	Ours	<b>0.7490</b>	<b>0.2498</b>	<b>0.0732</b>	<b>0.0383</b>	<b>0.0330</b>	<b>0.0276</b>	<b>0.0229</b>

Table 1: Quantitative comparison between a Graycode-based method and our method on two real-world HDR scenes. All metrics are error rates (↓ lower is better).

## 6 Reconstruction Accuracy under Varying Number of Patterns $N$

Increasing the number of patterns  $N$  improves reconstruction accuracy but also affects efficiency. To strike a balance between the two in learning-based methods, we conducted comprehensive experiments to determine the optimal value of  $N$ . We tested reconstruction accuracy for  $N$  ranging from 1 to 7 on planes and spheres, which is a widely adopted approach for precision calibration. The results are shown below:

Num of Patterns ( $N$ )	Plane	Sphere
1	1.10	1.18
2	0.85	0.91
<b>3</b>	<b>0.55</b>	<b>0.65</b>
4	0.55	0.57
5	0.50	0.52
6	0.41	0.49
7	0.40	0.47

Table 2: Evaluation of reconstruction errors (in mm) for different numbers of projected speckle patterns  $N$  on two representative surfaces. The optimal setting ( $N = 3$ ) achieves a good balance between accuracy and efficiency.

The results show that increasing  $N$  improves accuracy, but the improvement becomes marginal as  $N$  increases. When  $N = 3$ , we achieve accuracy comparable to using more frames, i.e.,  $N = 3$  is an elbow point that strikes an optimal balance between accuracy and efficiency.



## References

- [1] Prophesee. *Event Camera Evaluation Kit 4*, 2024. URL <https://www.prophesee.ai/event-camera-evk4/>. Accessed: Sep. 12, 2024.
- [2] Texas Instruments. *DLP6500 Digital Micromirror Device (DMD)*, 2023. URL <https://www.mouser.com/new/texas-instruments/ti-dlp6500-dmd/>. Accessed: Sep. 12, 2024.
- [3] Jiacheng Fu, Yueyi Zhang, Yue Li, Jiacheng Li, and Zhiwei Xiong. Fast 3d reconstruction via event-based structured light with spatio-temporal coding. *Optics Express*, 31(26):44588–44602, 2023.
- [4] Photoneo. Motioncam-3d m+. <https://www.photoneo.com/products/motioncam-3d-m-plus/>, 2024. Accessed: 2025-05-13.
- [5] Gangwei Xu, Yun Wang, Junda Cheng, Jinhui Tang, and Xin Yang. Accurate and efficient stereo matching via attention concatenation volume. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4):2461–2474, 2023.
- [6] Lahav Lipson, Zachary Teed, and Jia Deng. Raft-stereo: Multilevel recurrent field transforms for stereo matching. In *2021 International Conference on 3D Vision (3DV)*, pages 218–227. IEEE, 2021.
- [7] Xingyu Lu, Lei Sun, Diyang Gu, and Kaiwei Wang. Sge: structured light system based on gray code with an event camera. *Optics Express*, 32(26):46044–46061, 2024.