SpiderCam: Low-Power Snapshot Depth from Differential Defocus

Marcos A. Ferreira*,[1], Tianao Li*,[1], John Mamish*,[2], Josiah Hester[2], Yaman Sangar[2], Qi Guo[3], Emma Alexander[1]

[1]Northwestern University, [2]Georgia Institute of Technology, [3]Purdue University *Equal contributions.

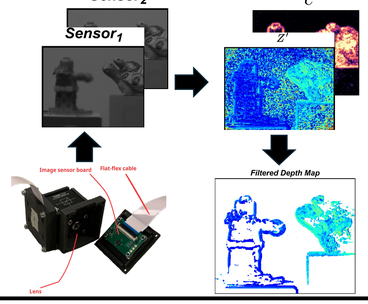
CVPR JUNE 3-7, 2026 DENVER COLORADO

Project page

Overview

We present the first sub-watt FPGA-based passive depth camera.

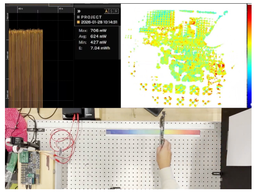
- **624 milliwatts** *power usage*
- **480 × 400** sparse *depth map*
- **31.5** frames per second
- **52** centimeters *working range*



Method

A beam splitter and dual low-power sensors generate a differential defocus signal for depth estimation. This signal is processed on an FPGA that is too small to contain a single image pair. Our design features:

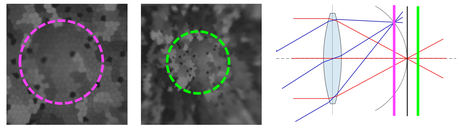
- Multiscale streaming architecture
- Compute-efficient kernels
- Hybrid fixed-/floating-point design
- Noise- and aberration-robust algorithm
- Radially varying optical parameters to compensate for Petzval field curvature

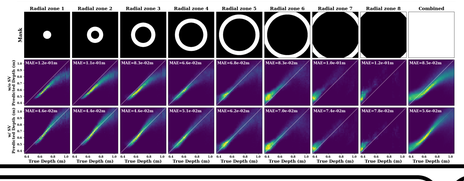


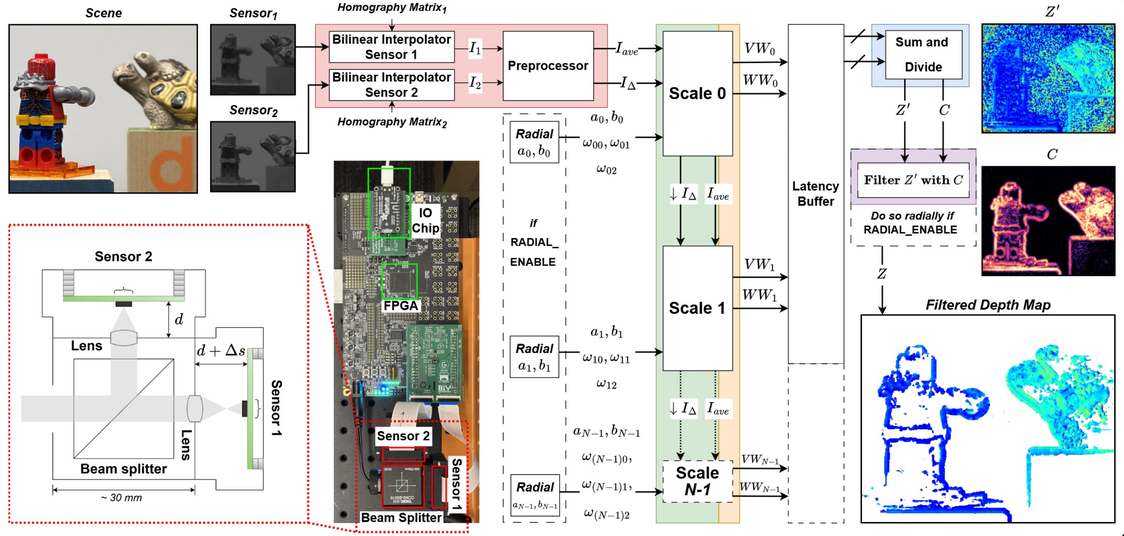
$$\text{Depth: } Z(x,y) = \frac{\sum_i \omega_i V_i(x,y) W_i(x,y)}{\sum_i \omega_i W_i^2(x,y)}$$

$$\text{Confidence: } C(x,y) = \sum_i \omega_i V_i(x,y) W_i(x,y)$$

$$V(x,y) = a\nabla^2 \tilde{I}(x,y), W(x,y) = bV(x,y) - \tilde{I}_\Delta(x,y)$$







Motivation

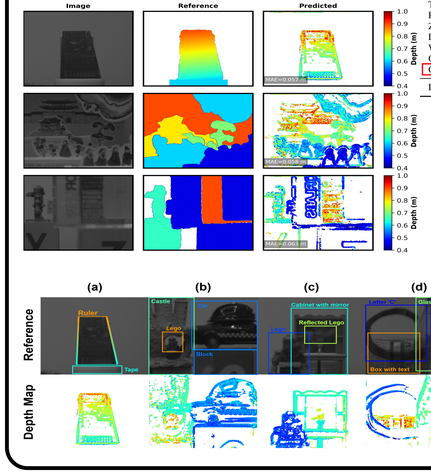
Previous work

- High power consumption
- Expensive optical setups
- Limited quantitative real-world evaluation

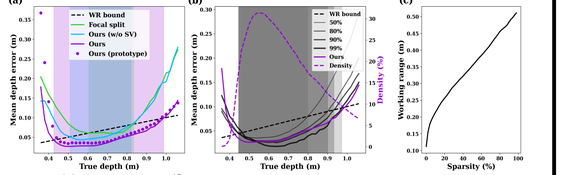
Our system

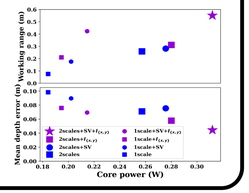
- State-of-the-art power efficiency on a resource-constrained FPGA
- Low-cost sensors, commercially available lens and 3D-printed enclosure
- Quantitatively validated on real sensor data

Results

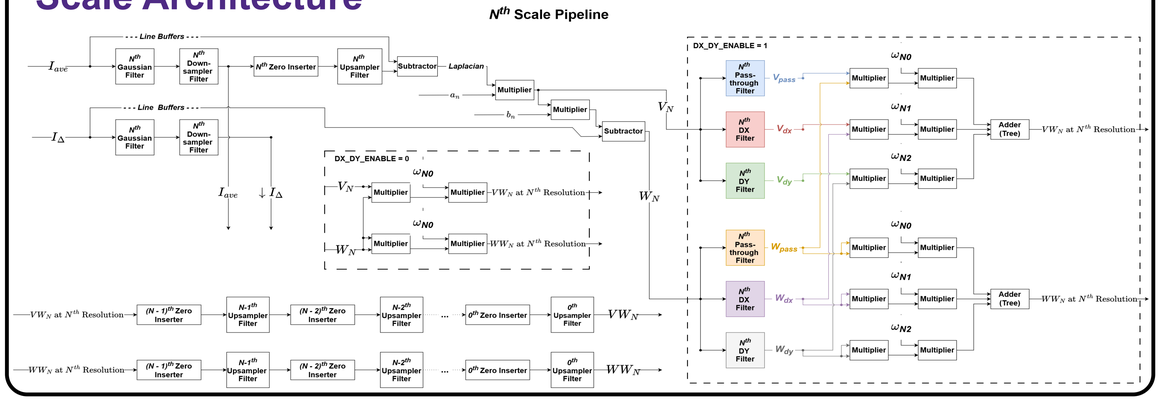


Name	Type	FPGA	Resolution	Core Power (W) Min	Core Power (W) Max	Real-World Quantitative Results: On-Device Working Range	Real-World Quantitative Results: Full System Power
Jin, 2014	Stereo	Kintex-6	640×480	1.93	2.67	none	none
Raj, 2014	DfD	Virtex-4[†]	400×400	0.46	0.58	0.77–0.80 m (**0.5% err**)	2 W + camera
Matteoccia, 2015	Stereo	Spartan-7	640×480	0.44	0.68	qualitative	2.5 W
Tfofis, 2015	Stereo	Kintex-7	1280×720	0.92	1.53	qualitative	2.8 W
Puglia, 2017	Stereo	Zynq Artix	1024×768	0.43	0.68	none	2 W
Zhang, 2018	Stereo	Kintex-7	1920×1080	0.89	1.23	none	none
Lu, 2021	Stereo	Kintex-7	1024×480	1.39	2.30	none	none
Wang, 2022	Stereo	Zynq US+	1920×1080	2.03	2.20	none	2.8 W + camera
Ours, normalized	DfD	Kintex-7	512×480	**0.42**	**0.55**	-	-
Ours, actual	DfD	ECP5	480×400	**0.24**	**0.31**	0.45–0.97 m (10% err)	**0.6 W**
Luo, 2025	DfD	N/A	480×360	N/A	N/A	**0.40–1.20 m** (10% err)	4.9 W





Scale Architecture



Relevant work

- Qi Guo, Emma Alexander, and Todd Zickler. Focal track: Depth and accommodation with oscillating lens deformation. In Proceedings of the IEEE international conference on computer vision, pages 966–974, 2017.
- Junjie Luo, John Mamish, Alan Fu, Thomas Concannon, Josiah Hester, Emma Alexander, and Qi Guo. Focal split: Untethered snapshot depth from differential defocus. In Proceedings of the Computer Vision and Pattern Recognition Conference, pages 26965–26974, 2025.