Live Demonstration: DVS-CIS Sensor Fusion System for Real-Time DNN-Based Object Detection

Mincheol Cha*, Keehyuk Lee*, Bobaro Chang*, Soosung Kim‡, Taeho Lee*,
Xuan Truong Nguyen*, Tae Sung Kim†, Hyuk-Jae Lee*, Hyunsurk Ryu*‡
*Dept. of Electrical and Computer Engineering, Seoul National University, South Korea
†Dept. of Electronic Engineering, Sun Moon University, South Korea
‡Neuro Reality Vision Corp., South Korea
{mccha, lkh099, bbrchang, taehov, truongnx, hjlee}@capp.snu.ac.kr, ts7kim@sunmoon.ac.kr, {kimsoosung, eric.ryu}@nrv.kr

Abstract—This demonstration presents a high-speed, energy-efficient sensor fusion system integrating CMOS Image Sensors (CIS) and Dynamic Vision Sensors (DVS) for advanced image recognition. Using CIS for high-res imaging and DVS for rapid event-driven capture, the FPGA-implemented architecture with an NPU running YOLOv3-Tiny achieves 18 ms inference latency with a minimal 2.78% mAP loss. Selective NPU activation based on DVS-detected regions yielded 31.5% power savings, while a custom receiver module efficiently fused DVS (13,900 fps) and CIS (60 fps) data. The system uses a power of 6.977 W on a Xilinx Zynq+ ZCU106 board.

I. INTRODUCTION

Dynamic vision sensors (DVS) [5] - an emerging technology - can be utilized in many tasks such as hand tracking [3] for various fields such as autonomous driving, surveillance, and robotics. Meanwhile, DNN accelerators such as [4] achieve high throughput and energy efficiency, appropriate for real-time surveillance systems. This work shows an energy-efficient daily surveillance system with DVS-CIS fusion and event-based NPU triggering. This demo is based on Submissions [1], [2] to ISCAS 2025.

II. DEMONSTRATION SETUP

Fig 1 illustrates a dual-FPGA sensor fusion system for real-time DVS and CIS data streams, supporting parallel processing and smooth integration. FPGA 1 manages data acquisition, buffering DVS and CIS streams in DDR4 memory for high throughput and low latency. Using an address decoder and inline transpose, it efficiently captures and organizes DVS data in DRAM. The data is then transferred via PCIe to the host PC for initial processing before moving to FPGA 2, where the NPU accelerates convolution operations, boosting performance. The final output returns to the PC for display, completing an efficient processing workflow. Fig 2 shows the setup for our real-time object detection system.

III. VISITOR EXPERIENCE

Visitors can experience real-time object detection with the proposed CIS-DVS fusion system. The demonstration highlights the performance distinctions between CIS-only and CIS-DVS combined systems, particularly in detecting small objects and operating effectively under low-light conditions.
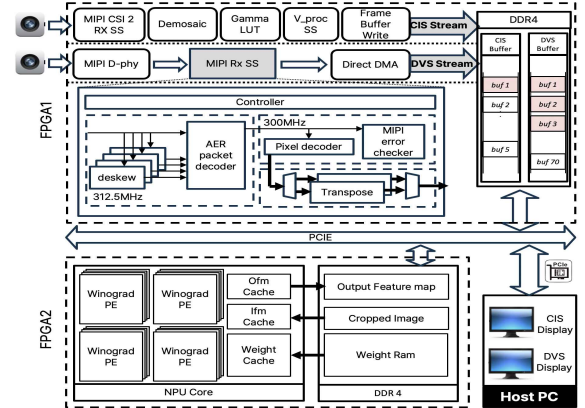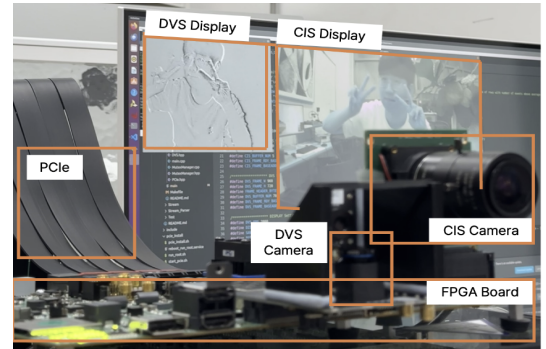

Fig. 1: The Proposed Architecture.

Fig. 2: Demonstration Setup. [2]

ACKNOWLEDGMENT

REFERENCES

[1] Mincheol Cha, Keehyuk Lee, Bobaro Chang, Soosung Kim, Taeho Lee Daniel Moon, Tae Sung Kim, Hyuk-Jae Lee, Hyunsurk Ryu, and Xuan Truong Nguyen. An Energy-Efficient Daily Surveillance System with DVS-CIS Sensor Fusion and Event-based NPU Triggering. In *Proceedings of the 2025 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2025.

[2] Mincheol Cha, Keehyuk Lee, Bobaro Chang, Soosung Kim, Taeho Lee, Xuan Truong Nguyen, Tae Sung Kim, and Hyunsurk Ryu. A DVS-CIS Sensor Data Receiver on FPGA with a 10 Gbps MIPI Controller. In *Proceedings of the 2025 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2025.

[3] Laura Duarte, Mohammad Safeea, and Pedro Neto. Event-based tracking of human hands. *Sensor Review*, 41(4):382–389, September 2021.

[4] Duy Thanh Nguyen, Tuan Nghia Nguyen, Hyun Kim, and Hyuk-Jae Lee. A High-Throughput and Power-Efficient FPGA Implementation of YOLO CNN for Object Detection. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 27(8):1861–1873, 2019.

[5] Bongki Son, Yunjae Suh, Sungho Kim, Heejae Jung, Jun-Seok Kim, Changwoo Shin, Keunju Park, Kyoobin Lee, Jinman Park, Jooyeon Woo, Yohan Roh, Hyunku Lee, Yibing Wang, Ilia Ovsiannikov, and Hyunsurk Ryu. 4.1 A 640×480 dynamic vision sensor with a 9μm pixel and 300Meps address-event representation. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 66–67, 2017.