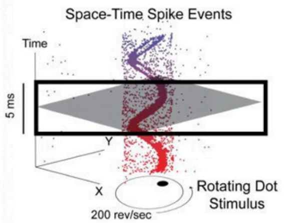
Samsung Research

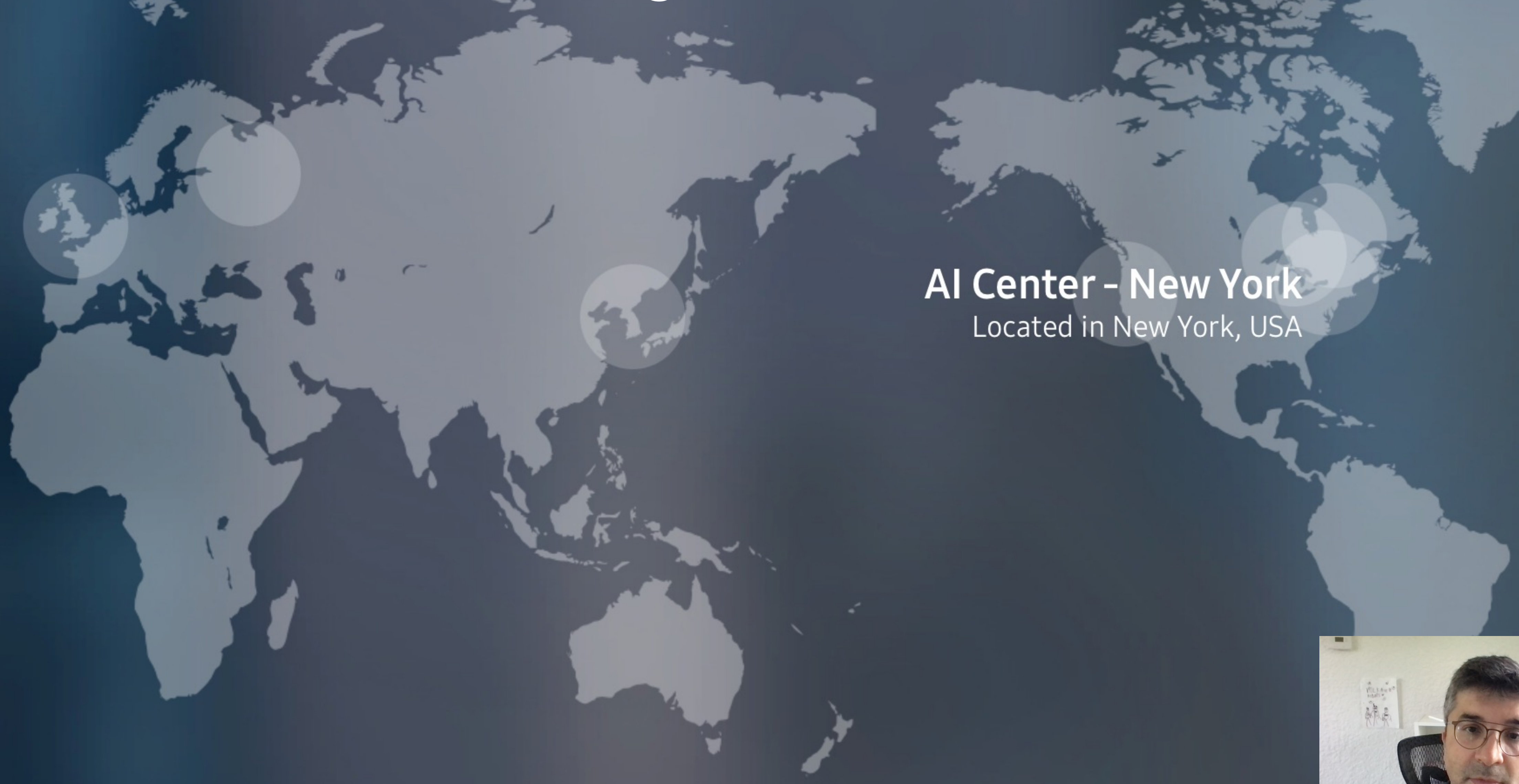


High Speed Perception-Action Systems with Event-Based Cameras

Anthony Bisulco, Daewon Lee,
Daniel D. Lee and Volkan Isler
Contact: a.bisulco@samsung.com

| June 19th, 2021 | Samsung Artificial Intelligence Center New York

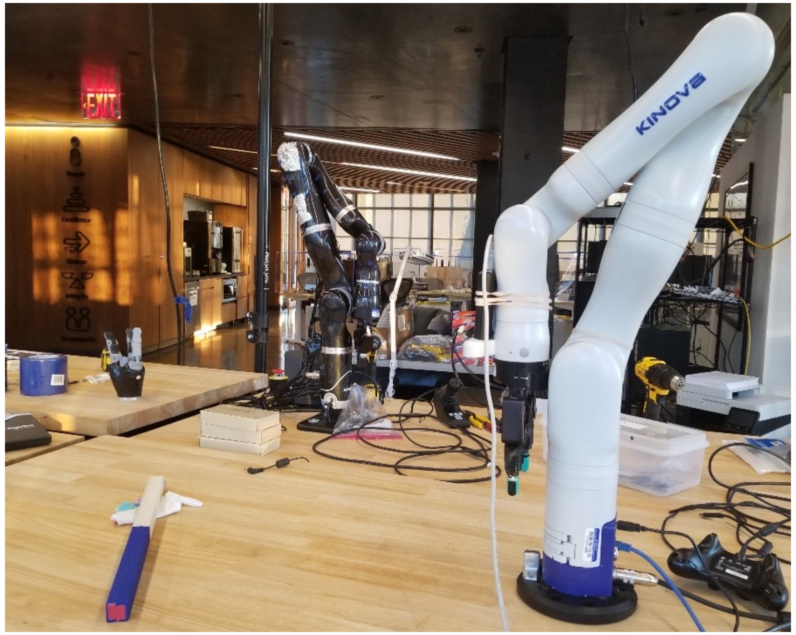
Samsung AI Centers

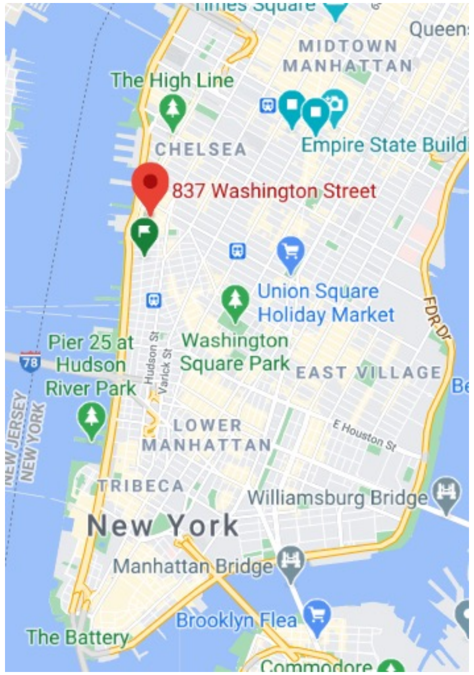


Samsung AI Center New York

Brief Introduction

SAIC-NY



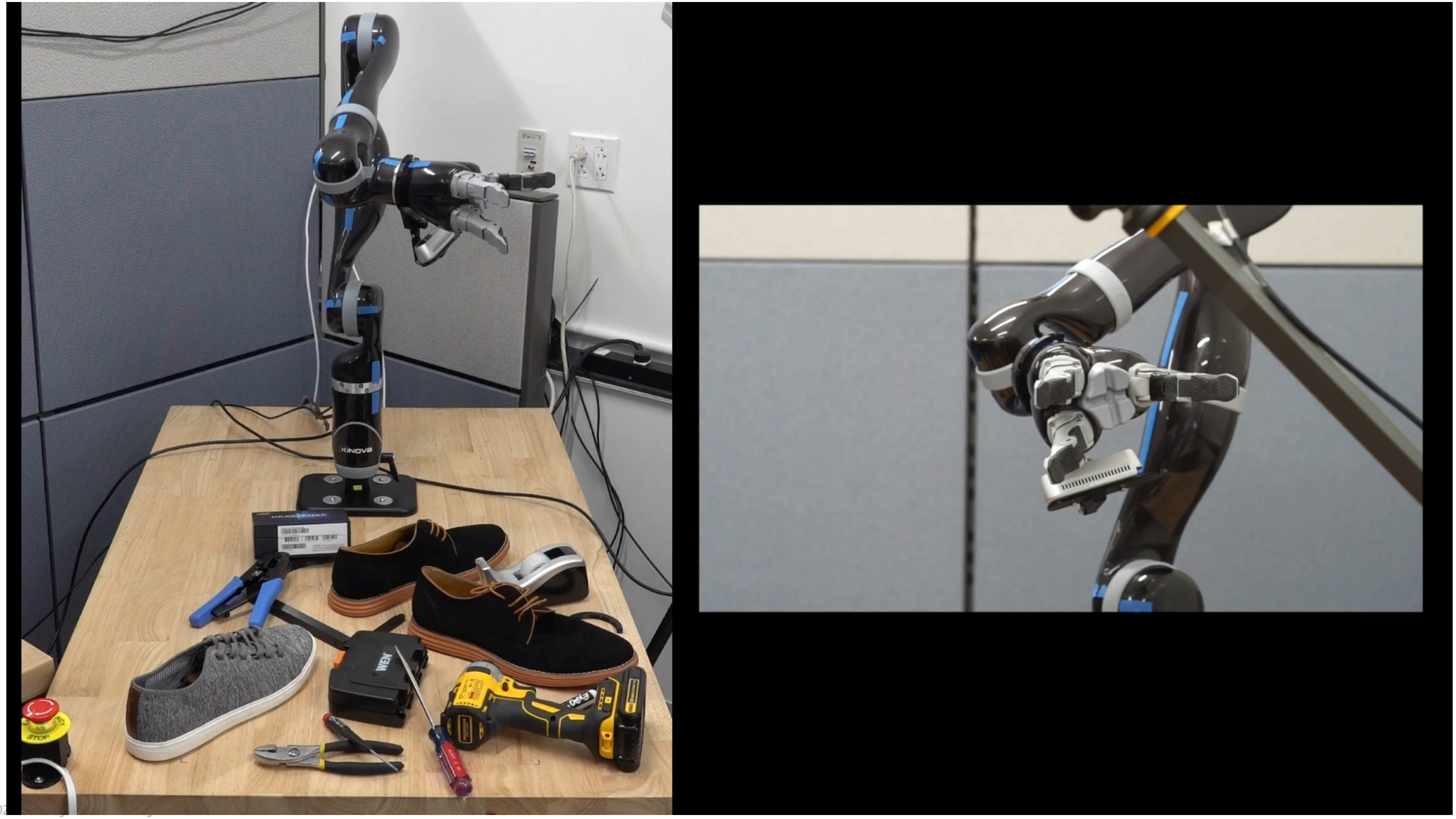


SAMSUNG

6th Floor
837 Washington St.
Chelsea

Samsung @ CES 2021

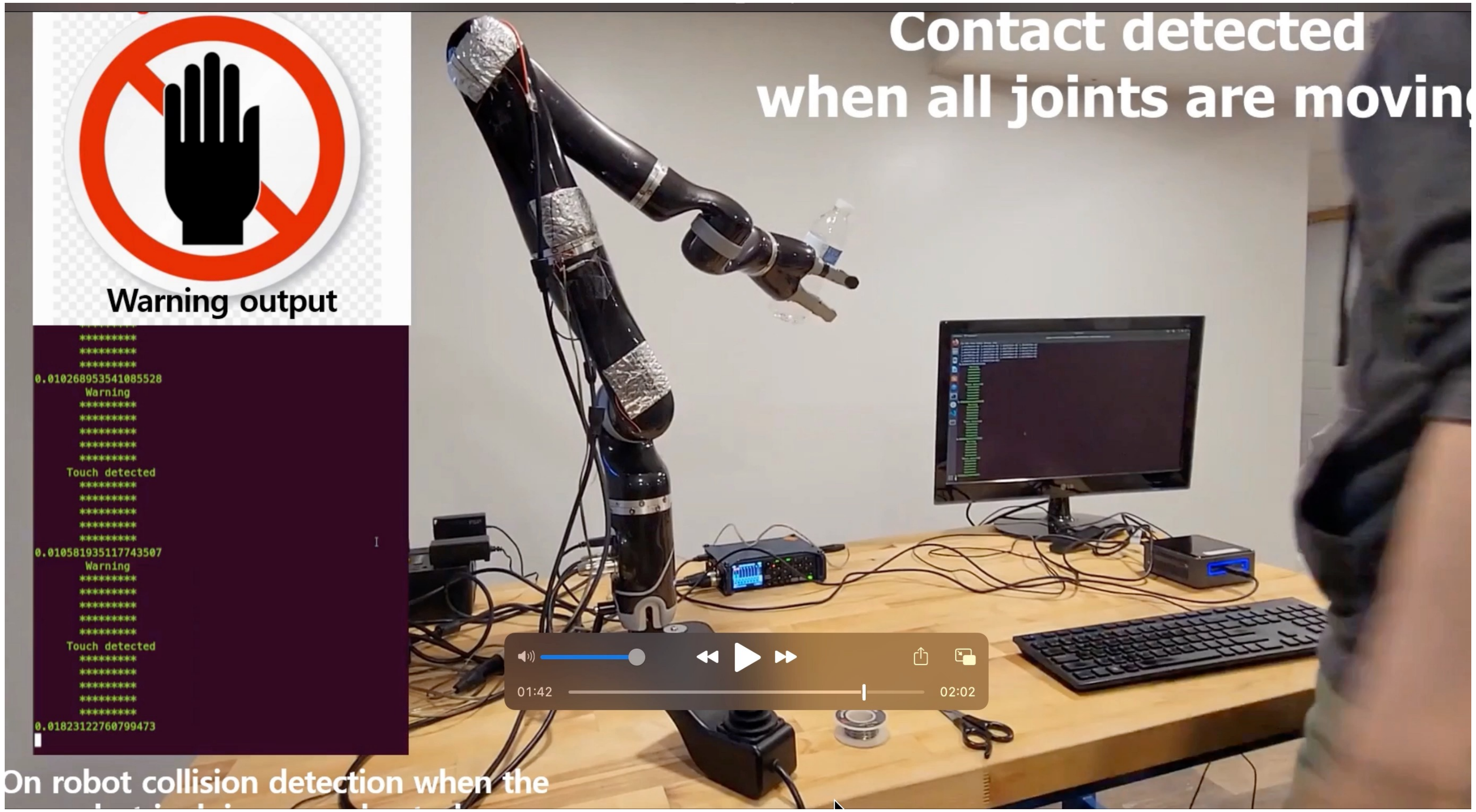
Manipulation in Clutter

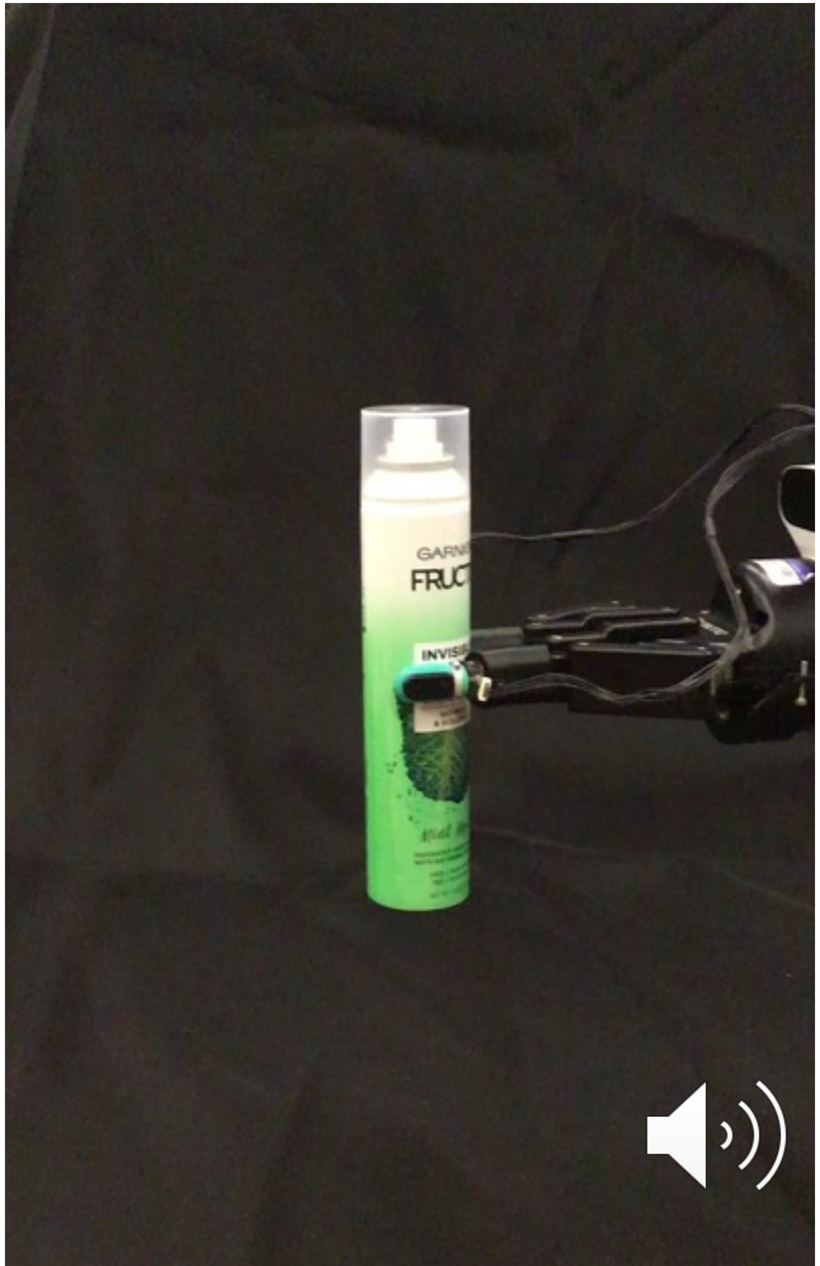



Grasping in clutter
CLOROX
Campbell's
Tomato
SOUP

Multimodal perception and control

Samsung Research





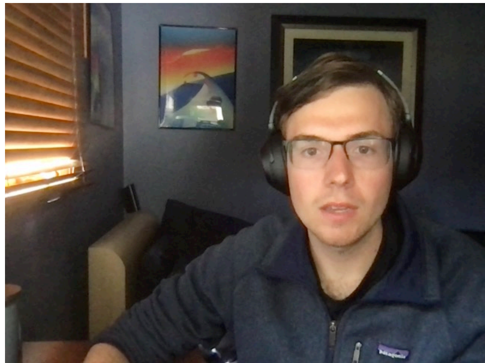
Samsung Research

Now on to DVS..

Outline

- **Event-Based Vision at SAIC-NY**
  - Efficient event-based classification(ICIP 2020)
  - Near-chip low bandwidth event-based classification(ISVLSI 2020)
  - Fast Motion Understanding(ICRA 2021)
  - Ongoing Work: Perception<->Action Systems

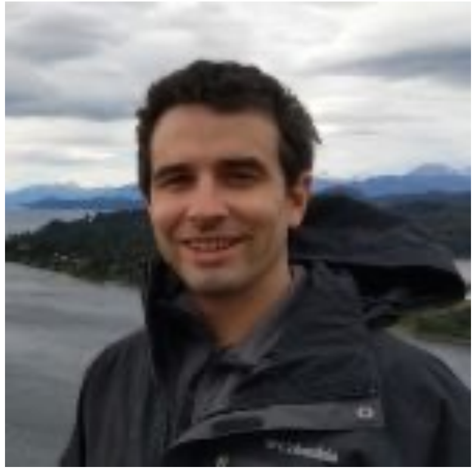
- **Conclusion**



On-Device Event Filtering with Binary Neural Networks for Pedestrian Detection Using Neuromorphic Vision Sensors

Fernando Cladera Ojeda, Anthony Bisulco, Daniel Kepple, Volkan Isler and Daniel D. Lee

2020 IEEE International Conference on Image Processing


Fernando Cladera


Anthony Bisulco

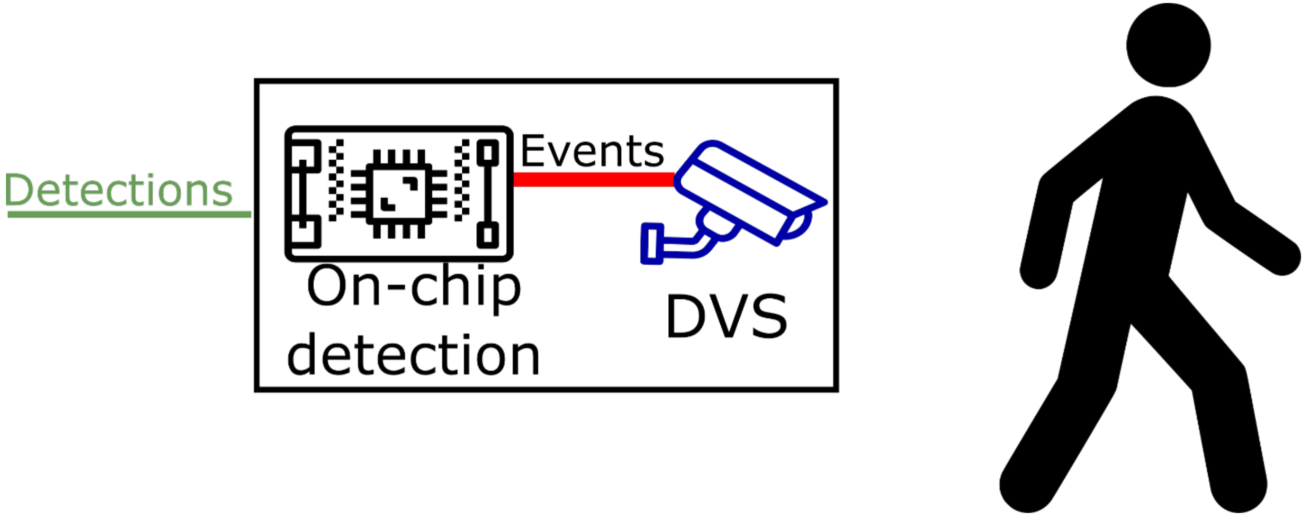

Daniel Kepple


Volkan Isler


Daniel D. Lee

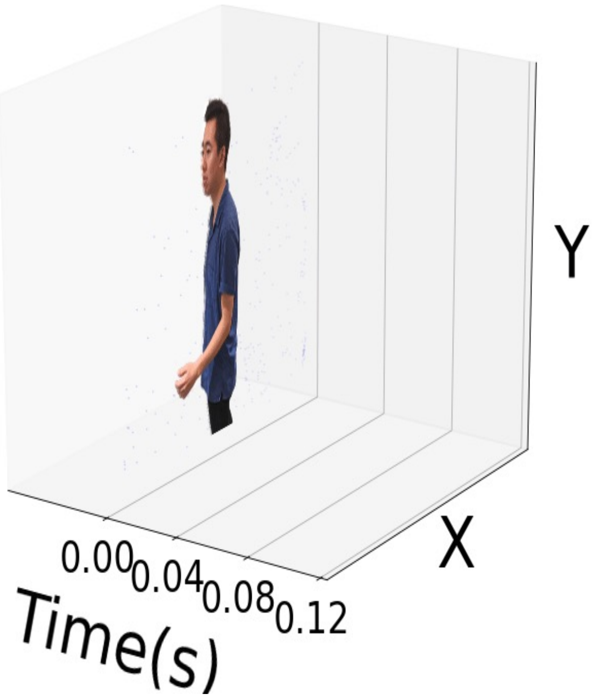
Problem: Always On Applications

- Pedestrian detection systems are **always-on** and **energy constrained**
- Traditional Convolutional Neural Networks require **compute-intensive** operations
- We present a low complexity architecture that **enables always-on event-based pedestrian detection**



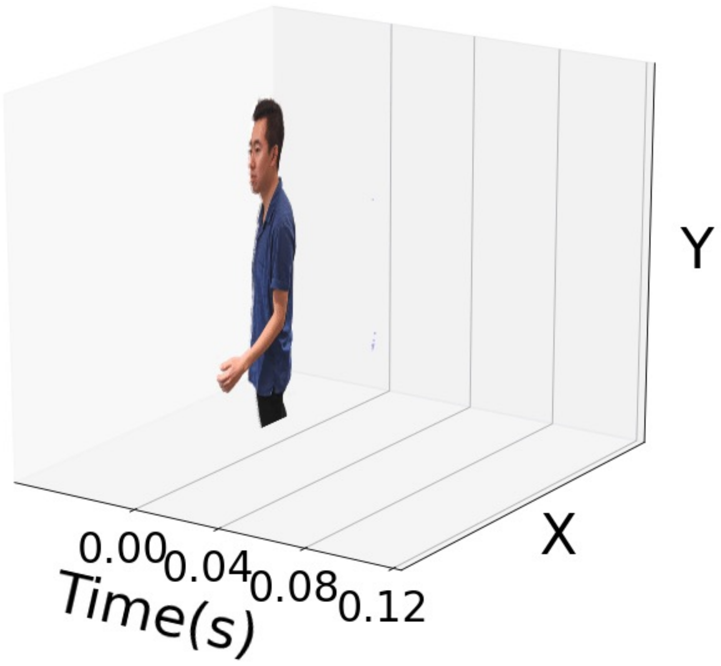
Our Method: DVS Detection Systems

Unfiltered Event Stream



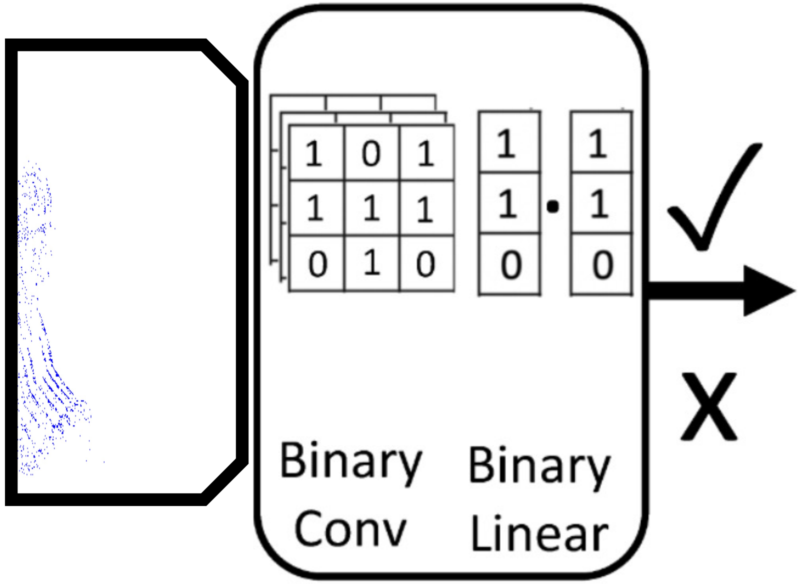
DVS Event Stream

Filtered Event Steam



Point Process Filter (PPF)

Windowed Binary Input

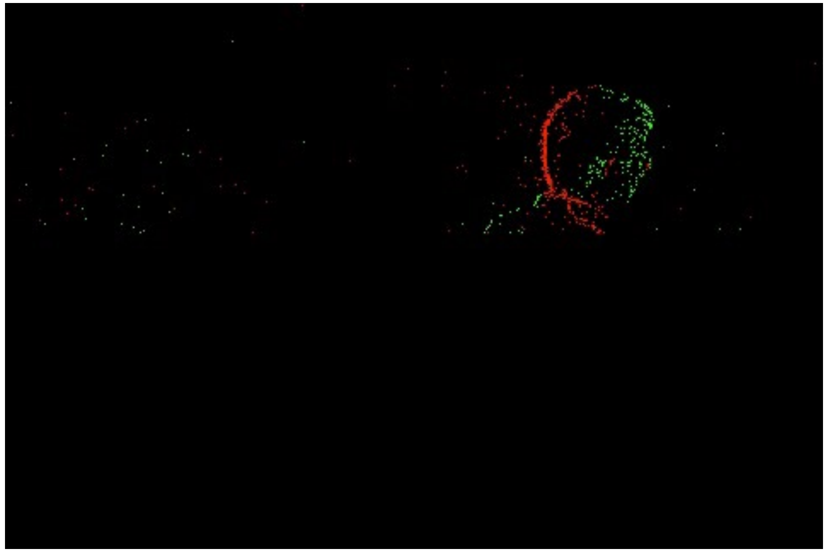


Binary Neural Network (BNN)

Pedestrian Dataset

- **Pedestrian dataset**: collected videos of people entering and leaving office
  - 273 ~2.5s clips of **pedestrian**
  - 548 ~0.75s clips of **negative examples**
- Negative examples: clothes, boxes, sticks and other visual stimuli
- Trained network on 80% tested on 20%

Positive Examples

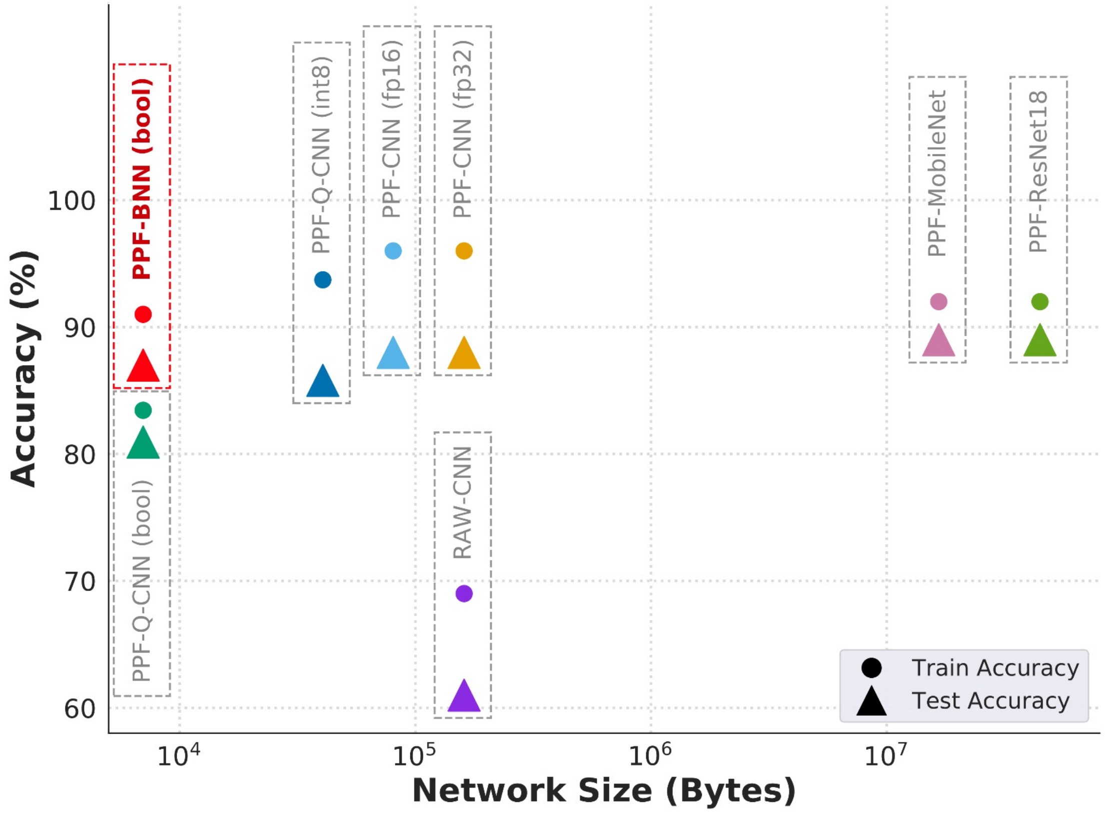


Negative Examples

Results

Samsung Research

- PPF
  - **Reduces noise** and **dimensionality** of the data
  - **Increases detection accuracy** by 23% compared to raw event stream
- Detection Module
  - Only **1% performance degradation** compared to PPF-CNN
  - **Lowest** network size, enabling embedded device applications



PPF: Point process filter
BNN: Inference with Binary Network
CNN: Inference using FP CNN
Q-CNN: Inference using quantized CNN

Conclusion

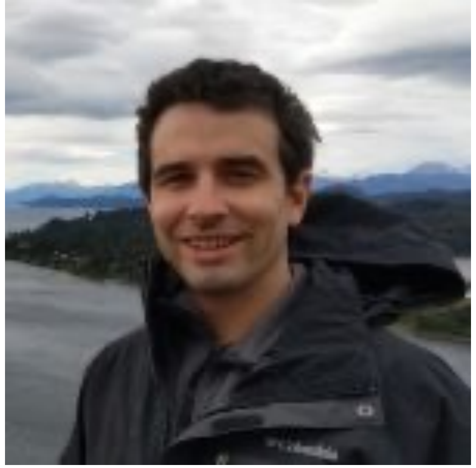
- **Low-complexity** object detection architecture for DVS:
  - PPF stage that boosts classification score by 23%
  - Detection module with low memory footprint
  - Enables DVS for low complexity environments: **home security and smart cities**
- **Future work:**
  - Near-Edge implementation

Near-chip Dynamic Vision Filtering for Low-Bandwidth Pedestrian Detection

Anthony Bisulco, Fernando Cladera Ojeda, Volkan Isler and Daniel D. Lee

2020 IEEE Computer Society Annual Symposium on VLSI

Anthony Bisulco



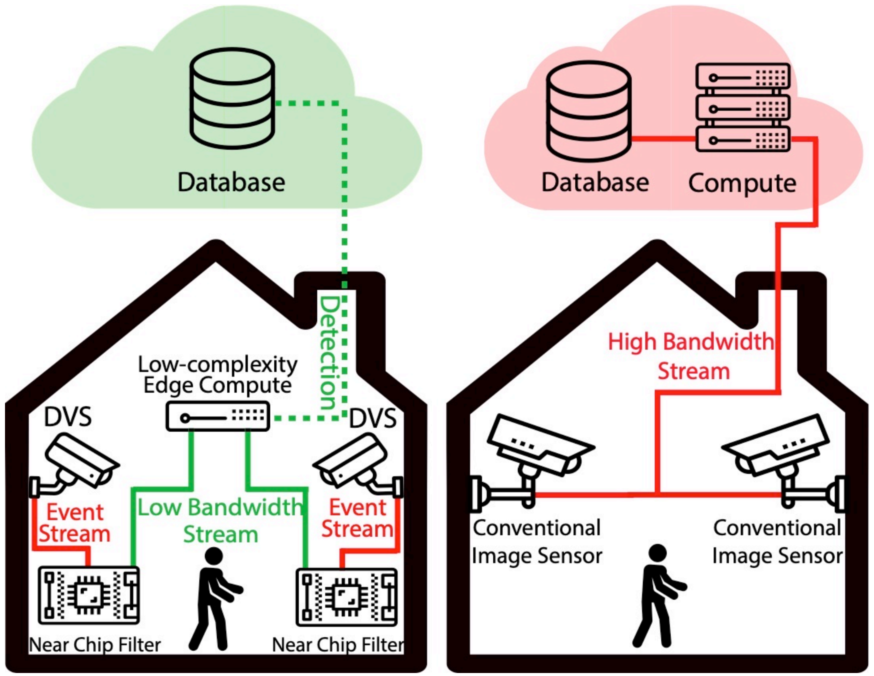
Fernando Cladera

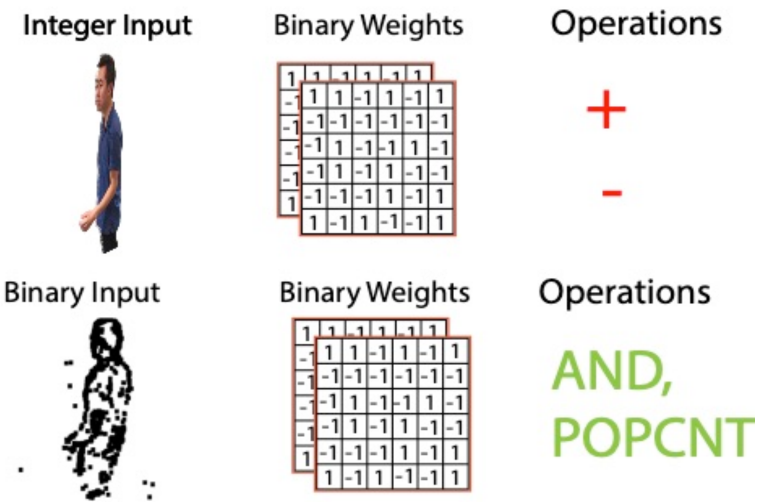
Volkan Isler

Daniel D. Lee

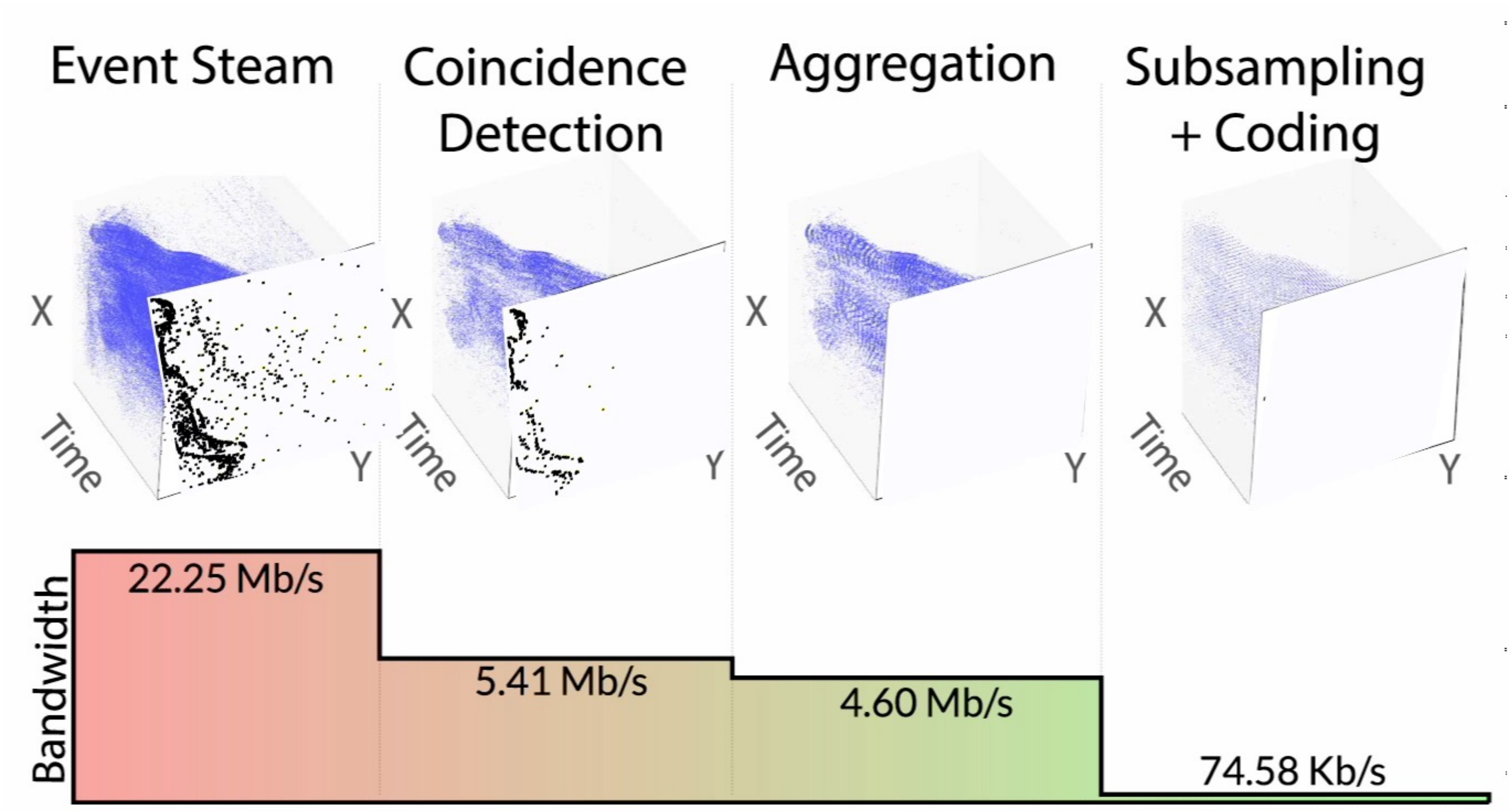
Problem: IoT Classification Systems

- **Goal:** DVS IoT Detection System
- **Problem**:
  - DVS high sampling rates lead to high bandwidths for IoT applications
  - IoT systems are low bandwidth
- **Solution:**
  - Develop on-chip algorithms for event-stream **compression**
  - Couple compressed stream with edge compute classification using Binary Neural Network



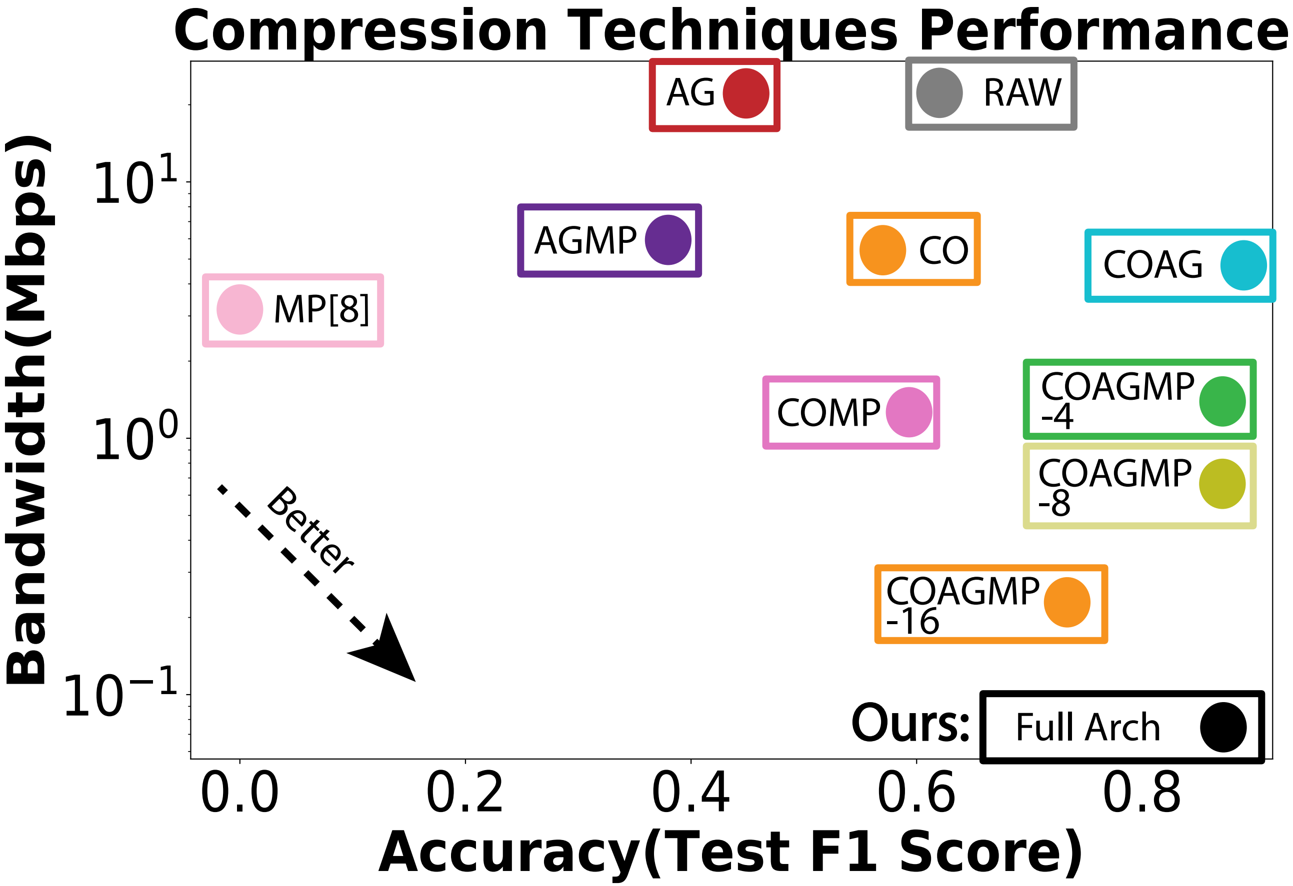


Our Method


Event Steam
Coincidence Detection
Aggregation
Subsampling + Coding
X
Time
Y
Bandwidth
22.25 Mb/s
5.41 Mb/s
4.60 Mb/s
74.58 Kb/s

Results

- Bandwidth reduction and increase of F1 score with few resources
  - Coincidence Detection/Max Pooling: **spatial reduction**
  - Aggregation/Huffman: **temporal reduction**
  - Full architecture:
    - **99.6% reduction** in bandwidth w.r.t. raw stream
    - **20% increase** in testing F1 score

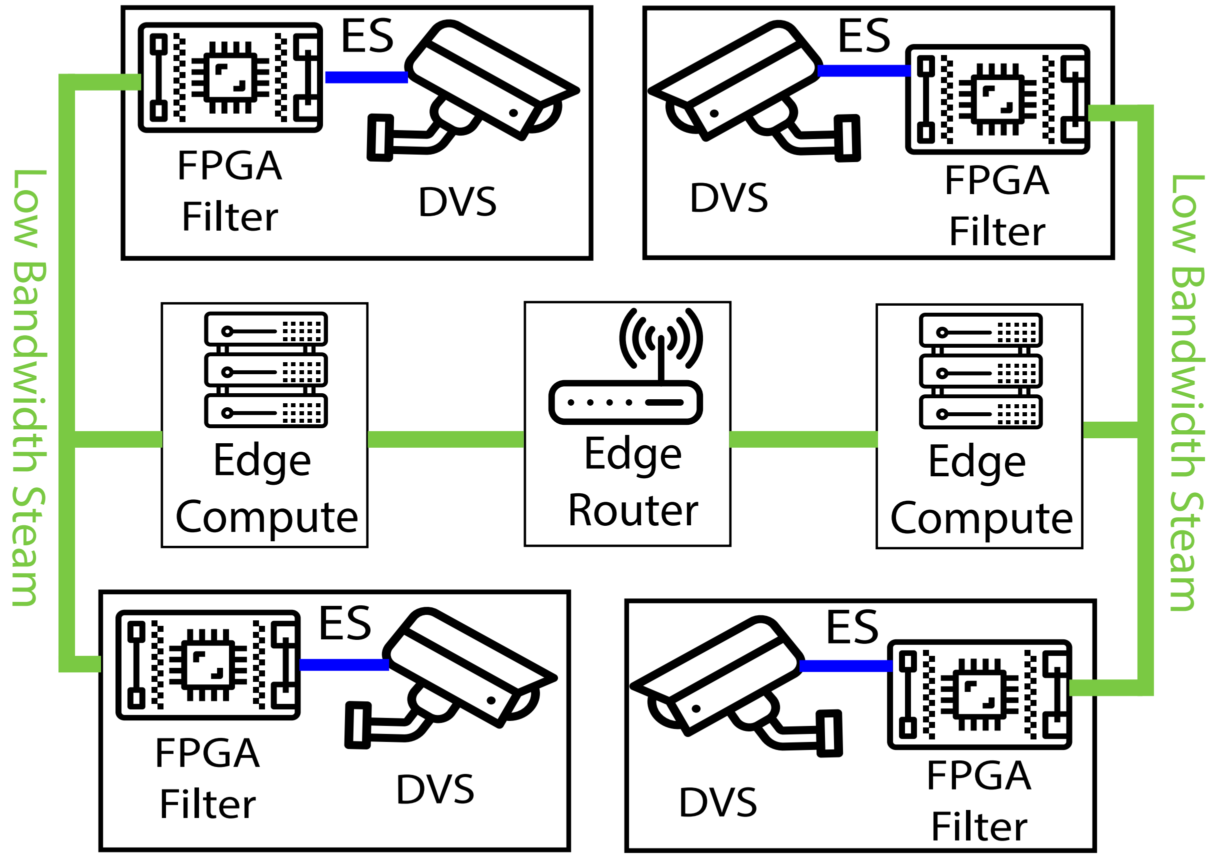


CO: Coincidence Detection
AG: Aggregation
MP-[#]: Max Pool (#,#)
Full Arch: **CoAgMP-8** + **Huffman**

Conclusion

- Filtering module: **reduces bandwidth** and **increases testing F1 score**
- Bandwidth reduction enables DVS usage on IoT networks
- Future work
  - Filtering module integration into DVS ASIC
  - Improvements in the compressor to further reduce bandwidth
  - Integrate multiple sensors to enhance detection performance

Sensor Network



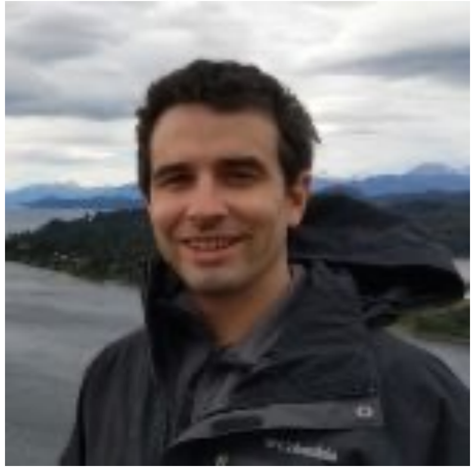
Fast Motion Understanding with Spatiotemporal Neural Networks and Dynamic Vision Sensors

Anthony Bisulco, Fernando Cladera Ojeda, Volkan Isler and Daniel D. Lee

2020 IEEE Computer Society Annual Symposium on VLSI

IEEE IGRA 2021

Anthony Bisulco



Fernando Cladera

Volkan Isler

Daniel D. Lee

Scenario

- Mobile robot systems need to quickly understand **rapid motion in dynamic environments**
  - Forests, Kitchens, Roads

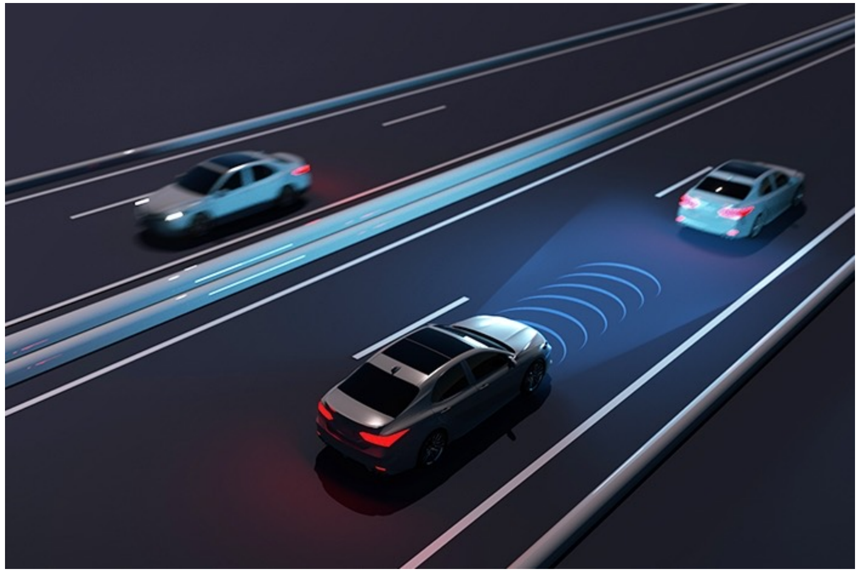
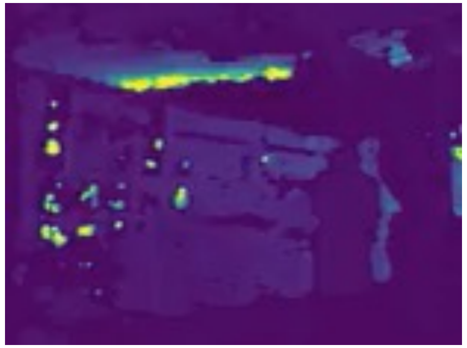


Image Courtesy of : https://phys.org/news/2020-10-treeswift-autonomous-robots-flight-forests.html, https://news.samsung.com/global/the-samsung-club-des-chefs-kitchen-heats-up-with-ai-assistance-at-ifa-2019

Problem: High-Speed Motion Understanding

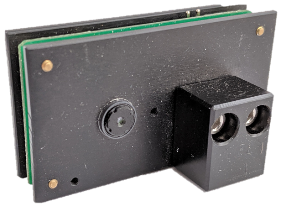
- Current sensor based-solutions are not suited for the energy and computational needs of micro mobile robot systems
    - **Active Sensors:** IR Depth Cameras, LIDAR, Radar
- **Dynamic vision sensors(DVS)** low latency sensing and energy consumption are attractive for mobile robot collision avoidance

Active Sensors Power Indicator

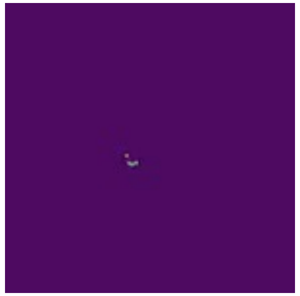


Depth Image

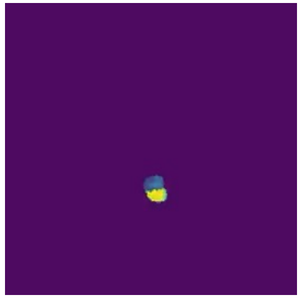
Passive Sensors Power Indicator



Dynamic Vision Sensor



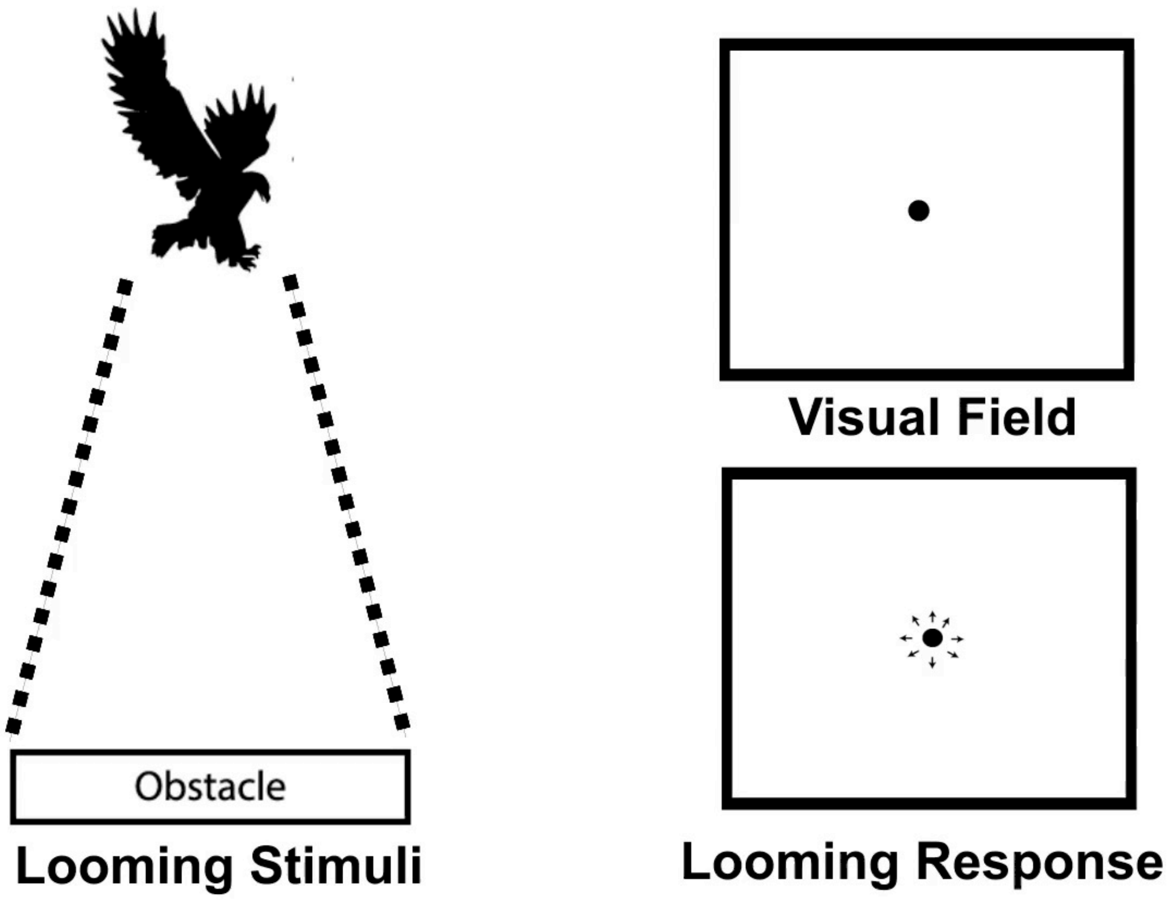
Event Video



Feature Tracking

Biological Motion Understanding

- Animals motion cues from looming response for avoidance [2][3]
  - Looming response is used in action planning for escape behavior
- Inspecting **spatiotemporal patterns** of the event stream could model these efficient avoidance maneuvers



Method

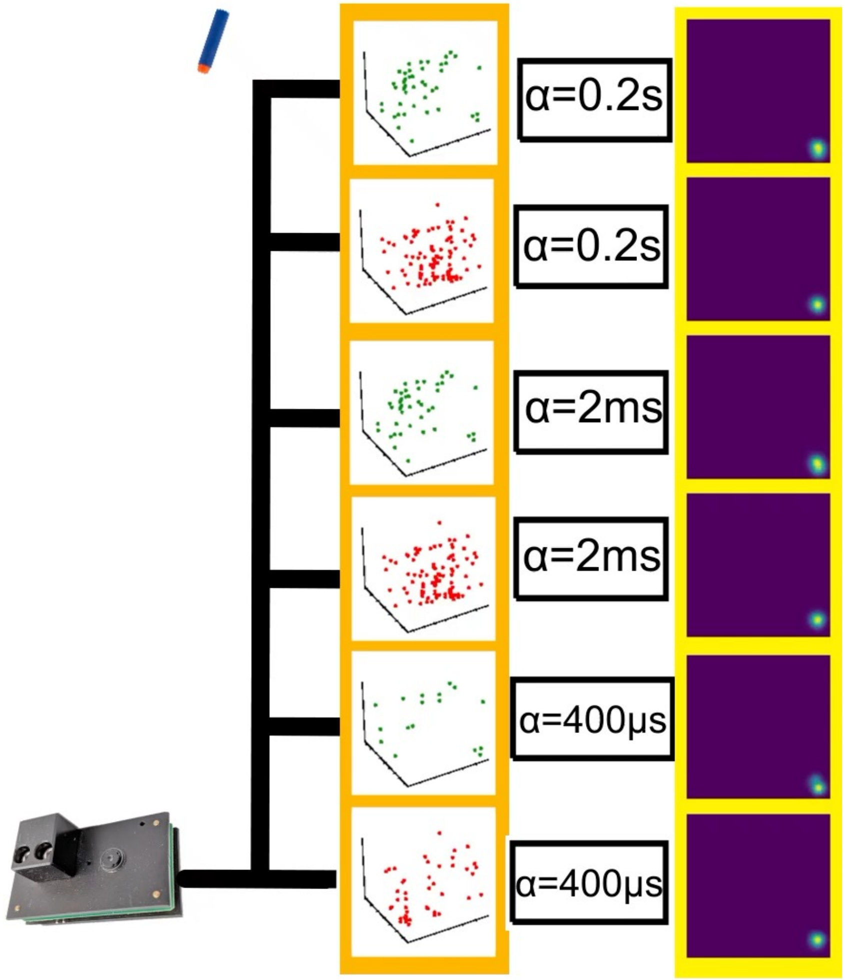
1. **Acquire**: Event-based camera outputs set of events (x, y, p, t)

2. **Encode:** History of events using bank of exponential filters per event polarity

$$y[n] = \alpha y[n-1] + (1-\alpha)x[n]$$

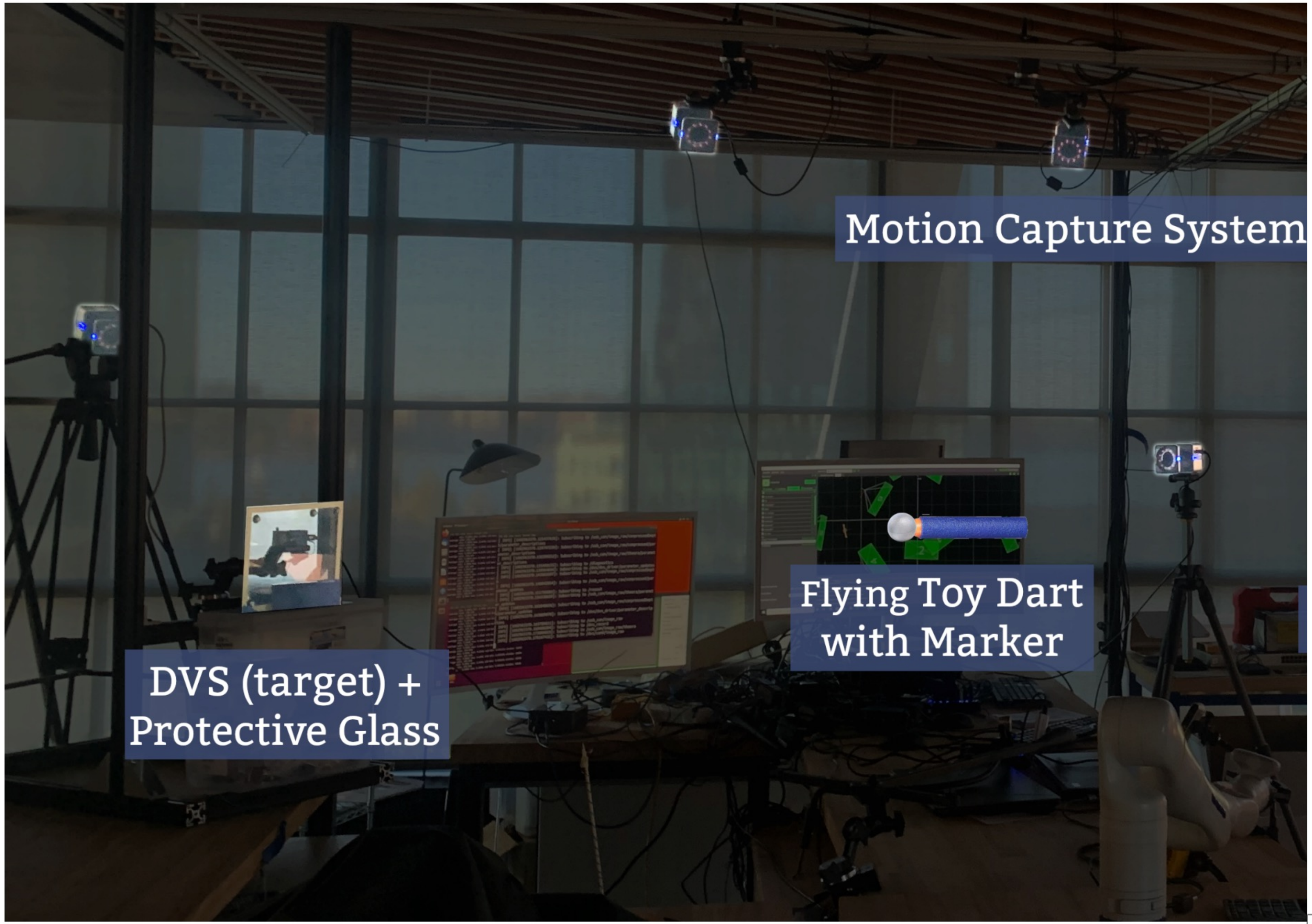
3. **Decode:** Extract Spatiotemporal features from filtered event set using CNN

4. **Estimate**: Time to Collision and Impact Location



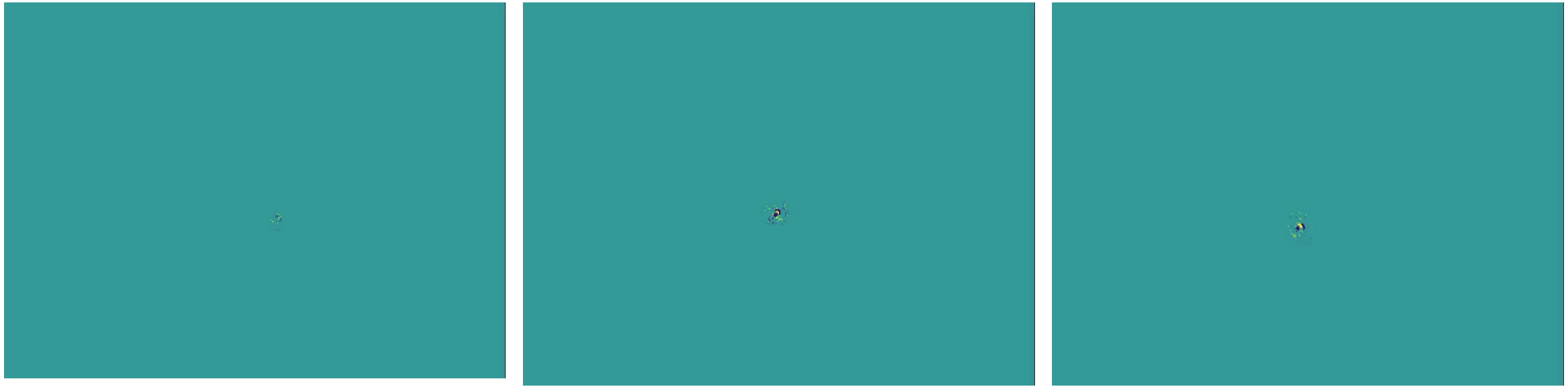
Method: Data Collection

- **Goal:** record the DVS events and object location while the object is approaching the sensor
- **Approach:**
  - Attach marker to object
  - Ball: Drop ball towards sensor
  - Dart: Flying dart



Method: Data Example

- **Toy Dart Info:**
  - Speed: 16-23m/s
  - Initial Range: 0.6m -1m
  - Time Range: 26ms -46ms
- **Ball Info**:
  - Speed: 1.2-4.8m/s
  - Initial Height: 0.4 m-1.2m
  - Time Range: 183ms-352ms
- Data: 240x240 unsigned 8-bit integer

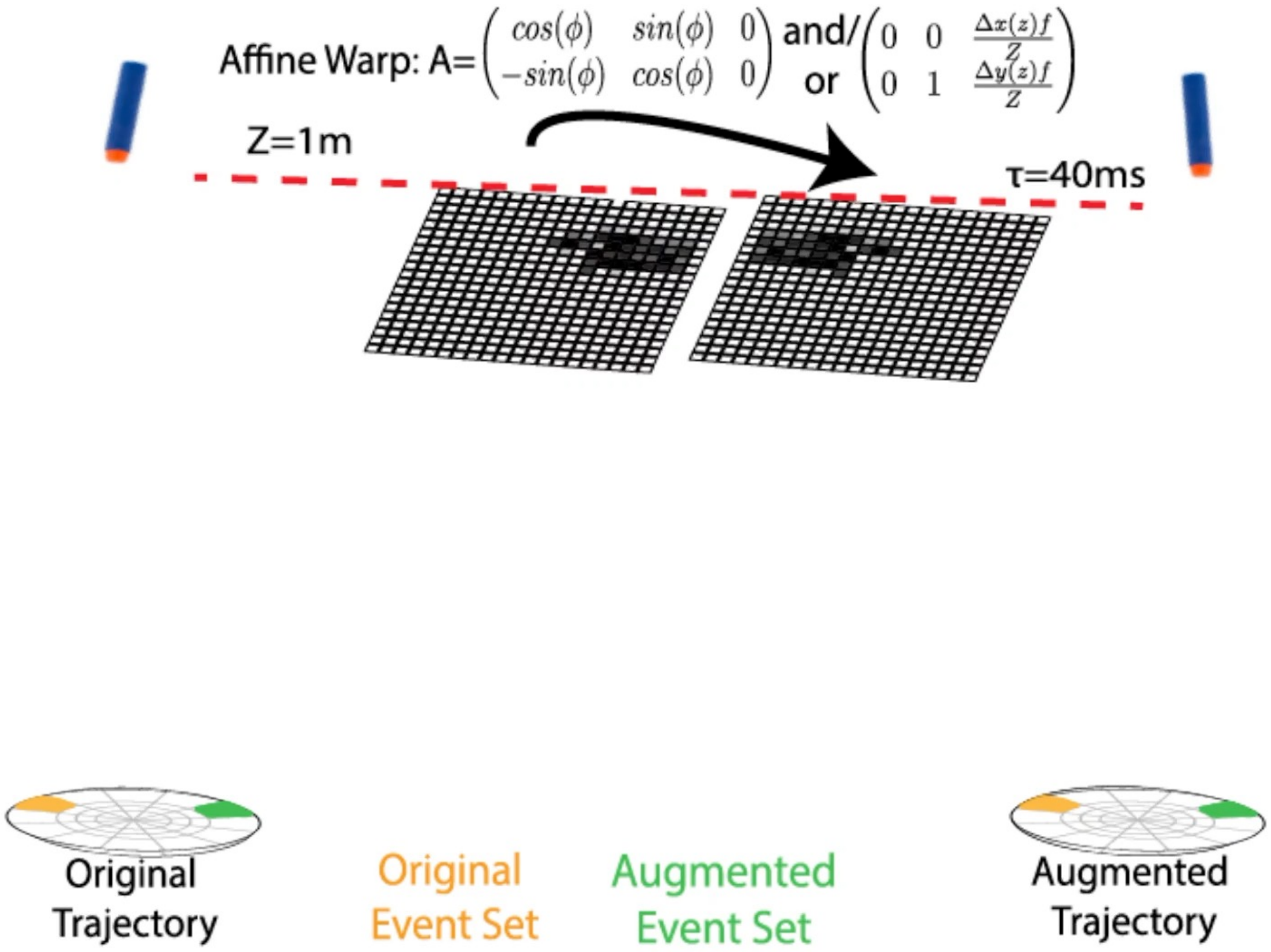


Toy Dart

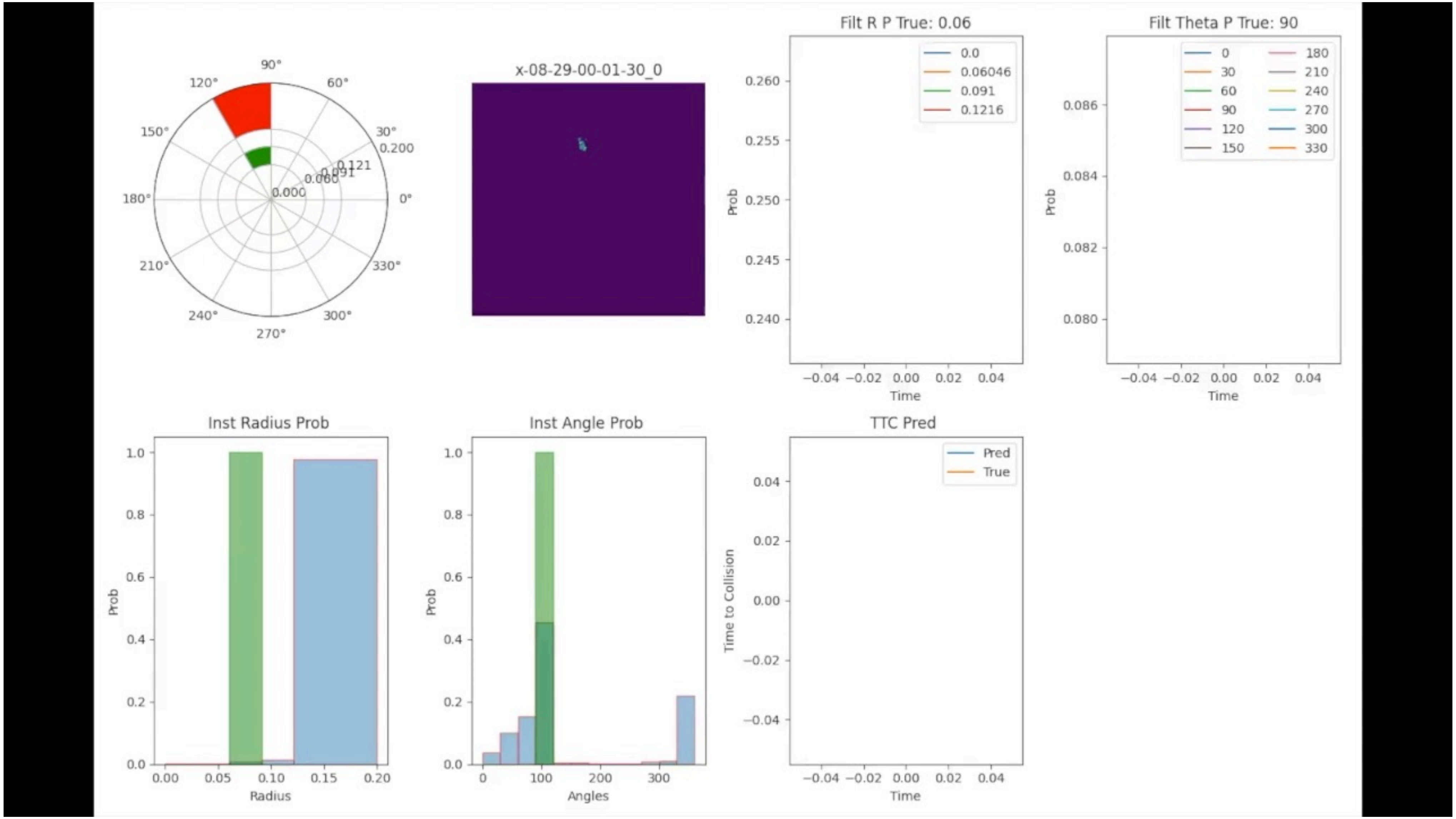
Ball

Method: Augmentation

- Data collection procedure can be time consuming and **expensive** for **limited samples**
- Developed augmentation procedure to increase dataset size from real-world samples
  - **Ball**: Perform set of static translations and rotations to event-data
  - **Toy-Dart**: Perform set of random rotations to event-data



Results: Example Fall



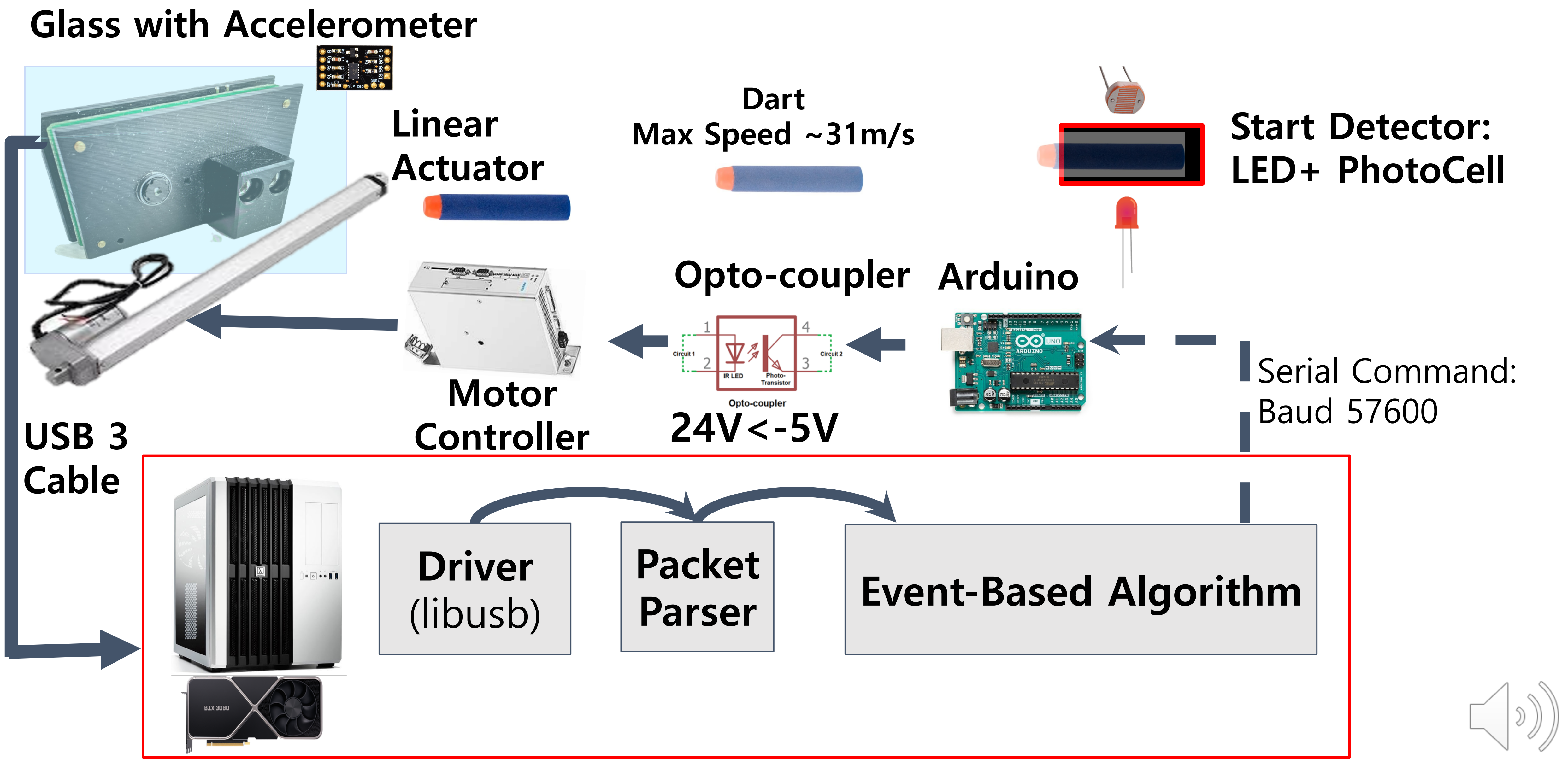
Conclusion

- We presented:
  - An efficient architecture for fast motion understanding using event based sensors
  - Data collection procedure
  - Data augmentation method to create additional real-world data for spatiotemporal model training
- **We showed that event-based sensors can be used to estimate the impact location and time to collision for fast moving objects (>20m/s) in mobile robotic applications.**
- Future Steps: Integrate the full system into a real-world setup

High Speed Perception-Action Systems with Event-Based Cameras

Ongoing

Perception<->Action Test Bed



Latencies

Samsung Research

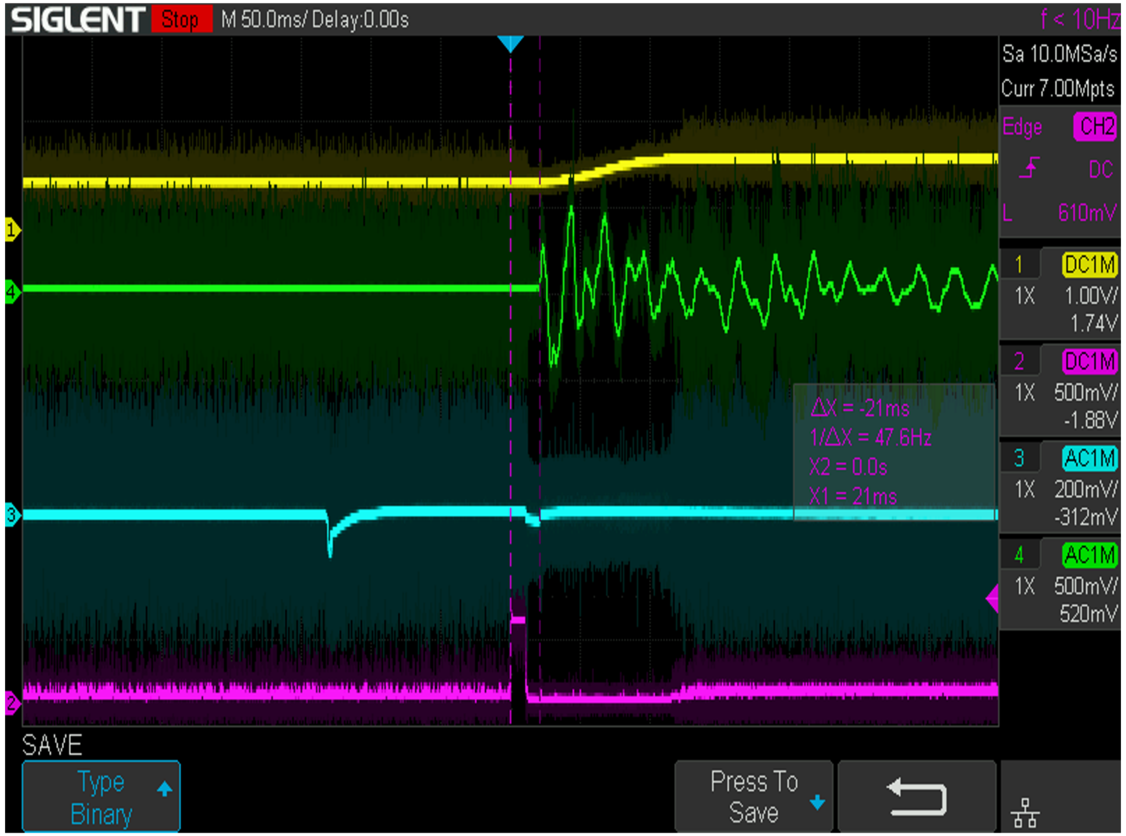
1. **Sense**
   a. **Observability Delay(~100ms):** Time press trigger-> Dart is minimally observable on th e camera(> 5 pixels wide) pixels Dart: ~12mm, 5m away shot
   b. **Communication Delay(~4ms):** Time between stimulus onset to pixel until the stimulus pixels is available in an array on the host, (FPGA Readout, USB Communication, Packet Decoding)
2. **Act**
   a. **Serial Latency(~1ms):** time from issuing the serial action command on the computer u ntil the motor receives this command
   b. **Action Latency(~33ms)**: time from 24V signal rising edge until reach end position in motion

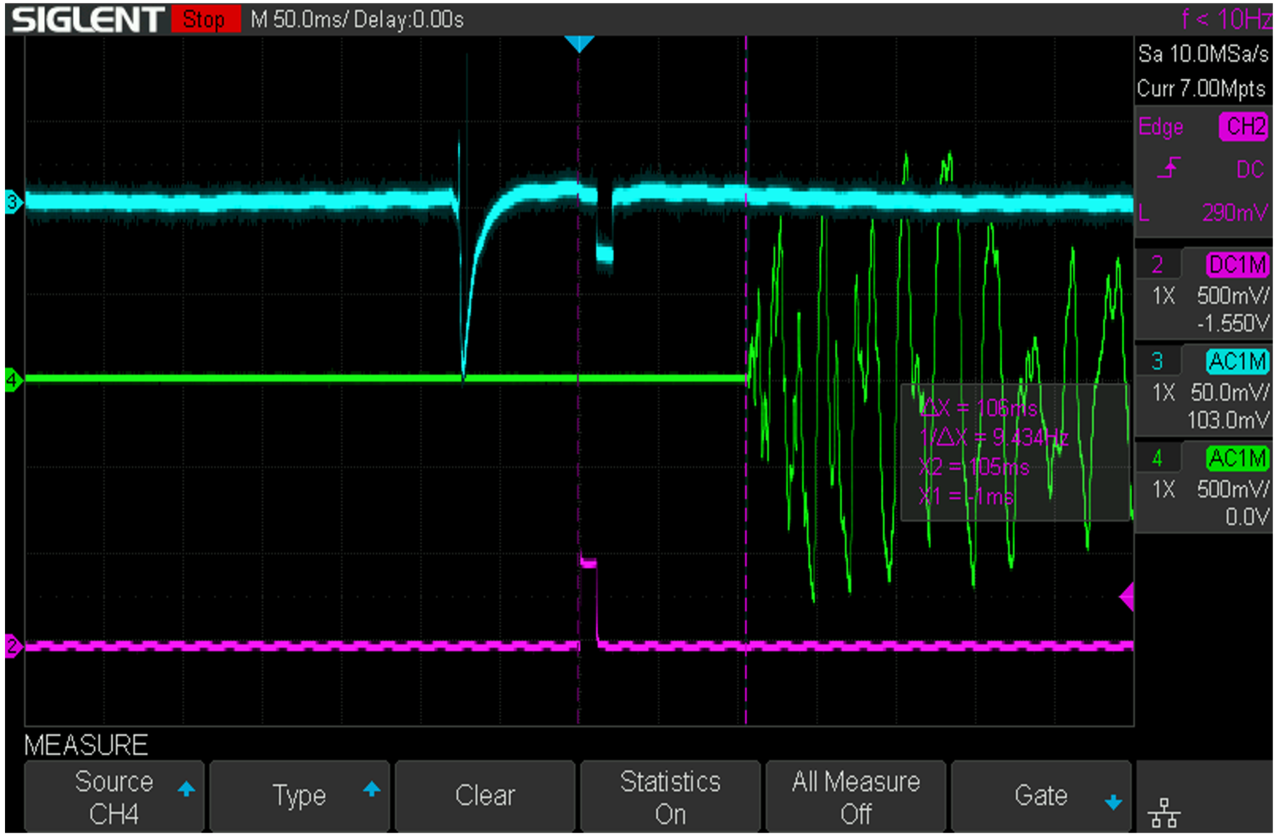
Observability Delay: Fix with Optics

Samsung Research

Result of Optics

Blue: Photocell(Start of Gun Shot)
Yellow: Motor Pos(Motion Encoder Position)
Green: Accelerometer(Collision with DVS)
Purple: Motion Trigger(Start of avoidance procedure)





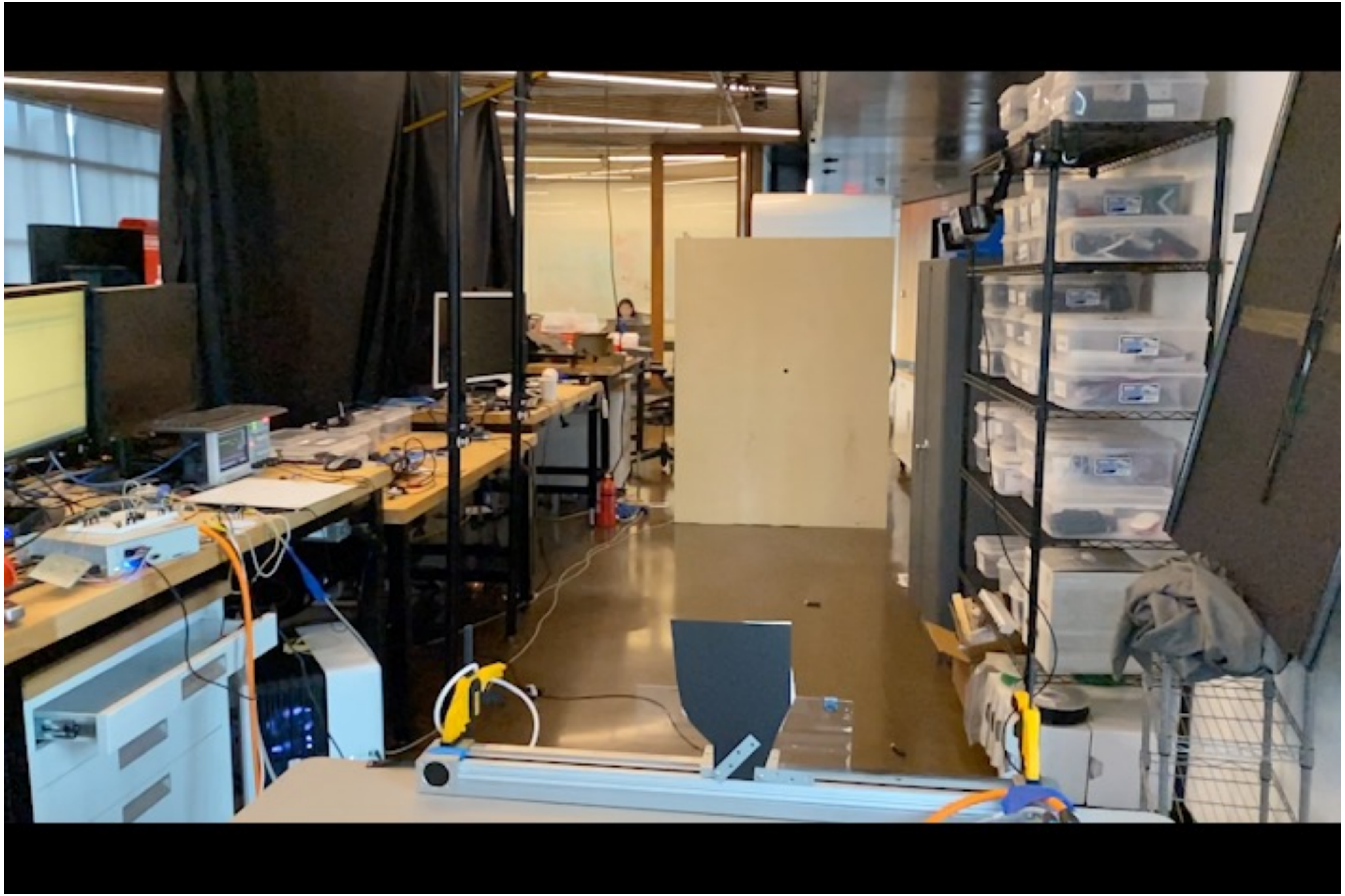
Time from Trigger->Collision: 21ms->106ms

Dodging Results: Real Time

Samsung Research

Slow Motion Dodge Video

Samsung Research



Conclusion

- Developed perception action systems and explored the system latencies present
  - Optics can help improve visibility of objects
  - Action latency remains large components in end-to-end system latencies
- Future Work
  - Improve our terminal state estimation system
  - Investigate other application of rapid state estimation
  - Evaluate predicative methods for event-based perception<->action
  - Unsupervised methods for perception

Samsung Research

Thank You

Questions: saic-ny@samsung.com