Industrial DVS Design; Key Features and Applications

Hyunsurk Eric Ryu, PhD
System LSI, Samsung Electronics

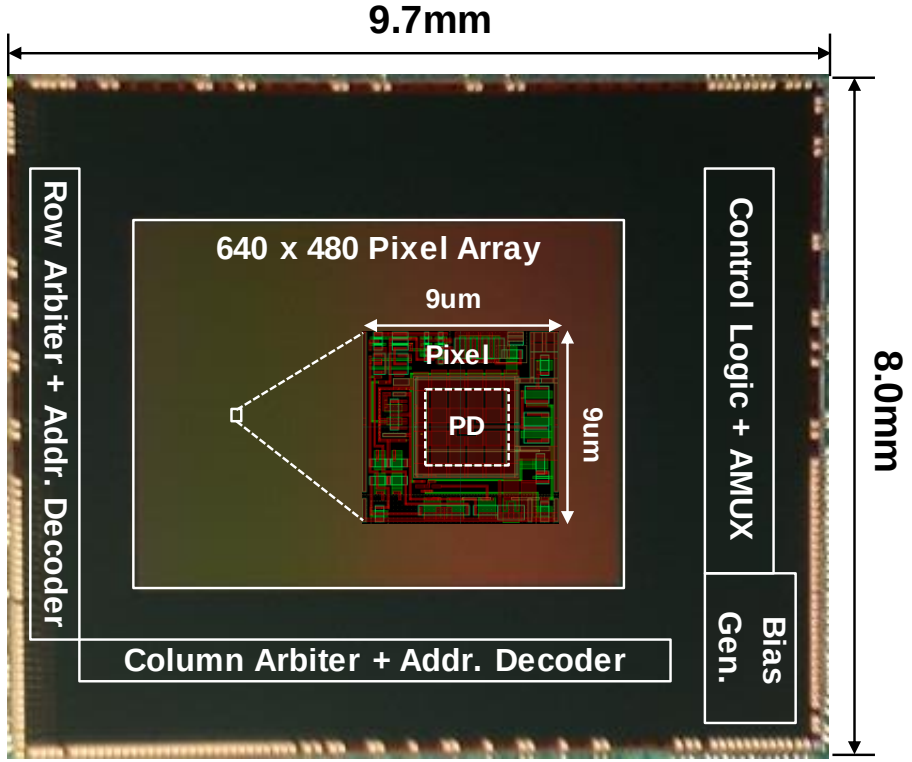
Outline

- Samsung S.LSI Dynamic Vision Sensor
- Key Features
  - Low Latency
  - Minimized Motion Artifacts
  - Anti-Flicker
  - Low Power Operation
- Applications
  - Sparse Edge-based Object Recognition
  - Pose Estimation with Visual Information: DVS-*SLAM

*SLAM: Simultaneous Localization and Mapping

- Summary and Discussion

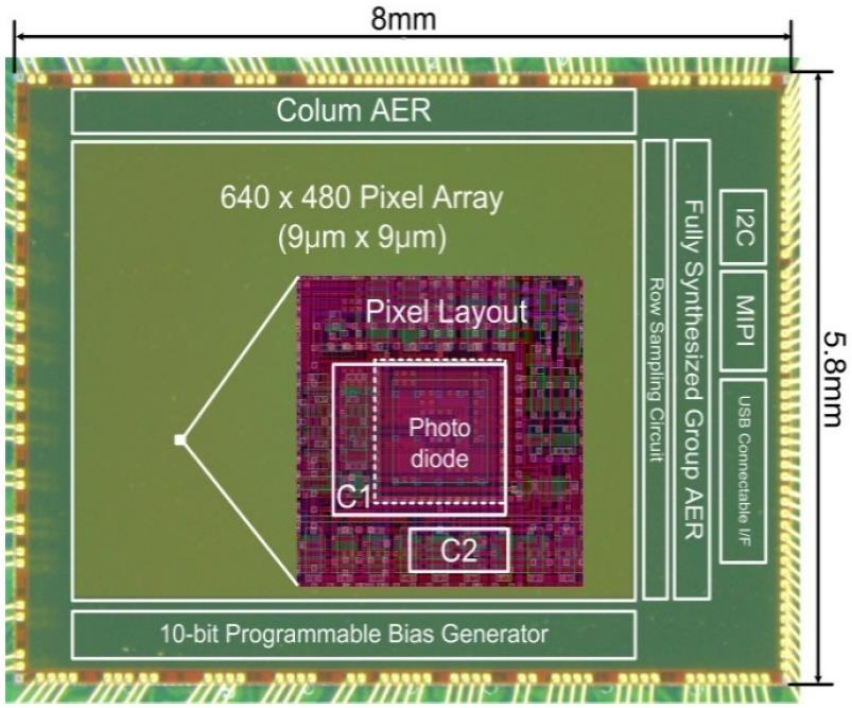
DVS Gen1 (R&D Ver., 2014)



- Brief Specifications
  - 640 x 480 Pixel Array with 1/2.5-inch optics
  - Pixel size : 9 um
  - Dynamic range : 66 dB (5~10,000 Lux)
  - Max. event processing rate : 6.5 Meps* **eps: event per second*
  - Minimum detectable contrast (50% response) : < 19%
  - Interface : 20-bit Parallel
  - Typical power consumption : 15 mW
- Key Features
  - Small pixel size (9 um)

- Full custom logic design
- Fabricated using the Samsung 90-nm Back Side Illumination (BSI) CIS process

DVS Gen2 (R&D Ver., 2016)



Chip Photograph

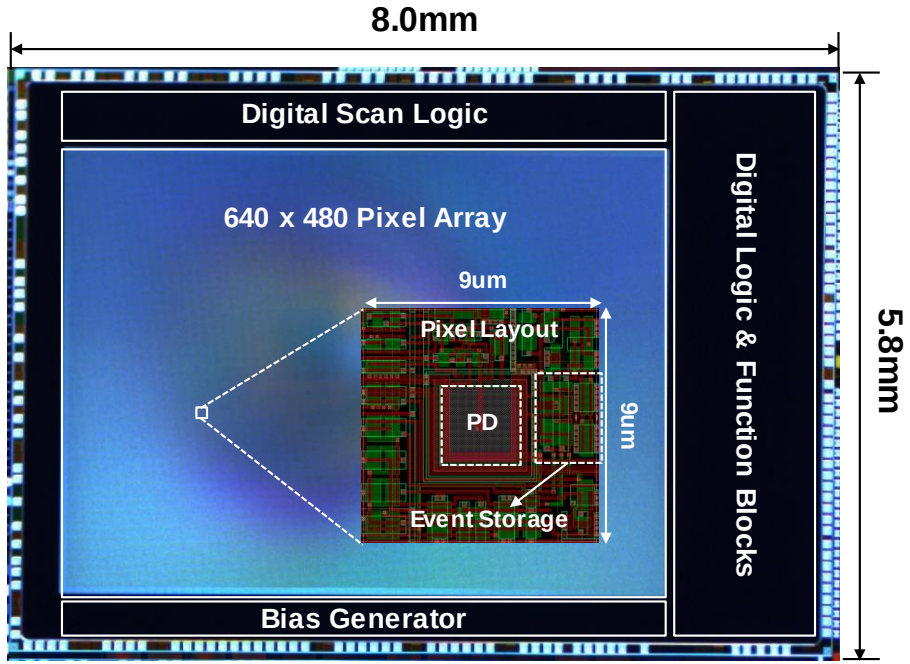
(ISSCC '17)

- Brief Specifications
  - 640 x 480 Pixel Array with 1/2.5-inch optics
  - Pixel size : 9 um
  - Dynamic range : 90 dB (3~100,000 Lux)
  - Max. event processing rate : 300 Meps
  - Minimum detectable contrast (50% response) : < 19%
  - Interface : MIPI(1Gbps 4-lane), USB connectable Parallel, I^2C
  - Typical power consumption : 80 mW
- Key Features
  - Standard digital I/F IP supported
  - Digitally synthesized G-AER* for high throughput

**G-AER: Group Address Event Representation*

- High data throughput using digitally synthesized G-AER scheme

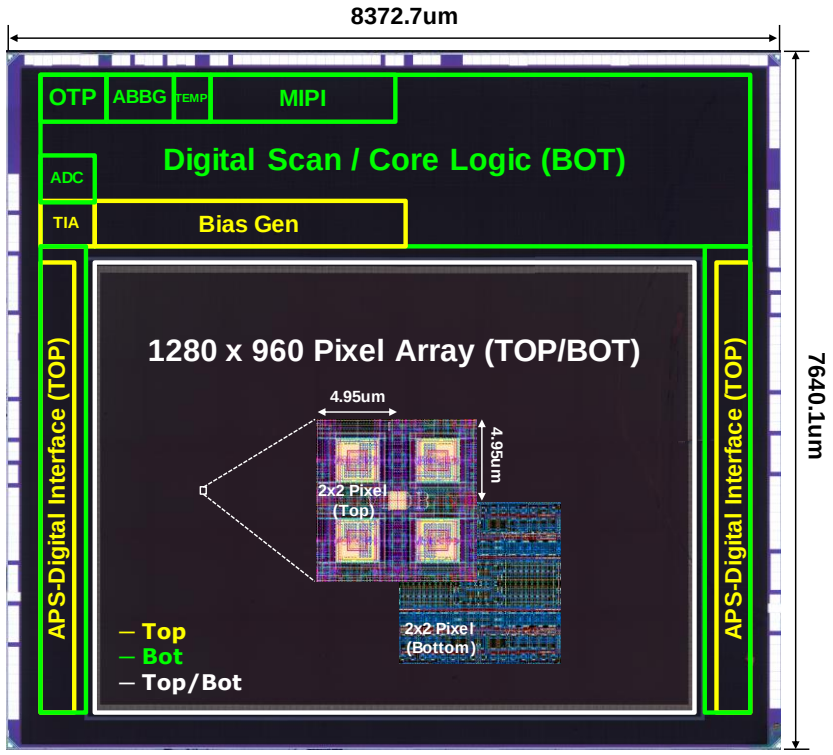
DVS Gen3 (Product Ver., 2018)



- Brief Specifications
  - 640 x 480 Pixel Array with 1/2.5-inch optics
  - Pixel size : 9 um
  - Dynamic range : 90 dB (3~100,000 Lux)
  - Effective frame rate : > 2,000 fps (MIPI)
  - Minimum detectable contrast (99.9% response) : < 27.5%
  - Interface : MIPI(1Gbps 4-lane), USB connectable Parallel, I^2C
  - Typical power consumption : 65 mW
- Key Features
  - Global hold, Global reset, Column scan readout

- Motion image artifact minimized by using global hold and global reset
- Timestamp error minimized by applying sequential column scan readout

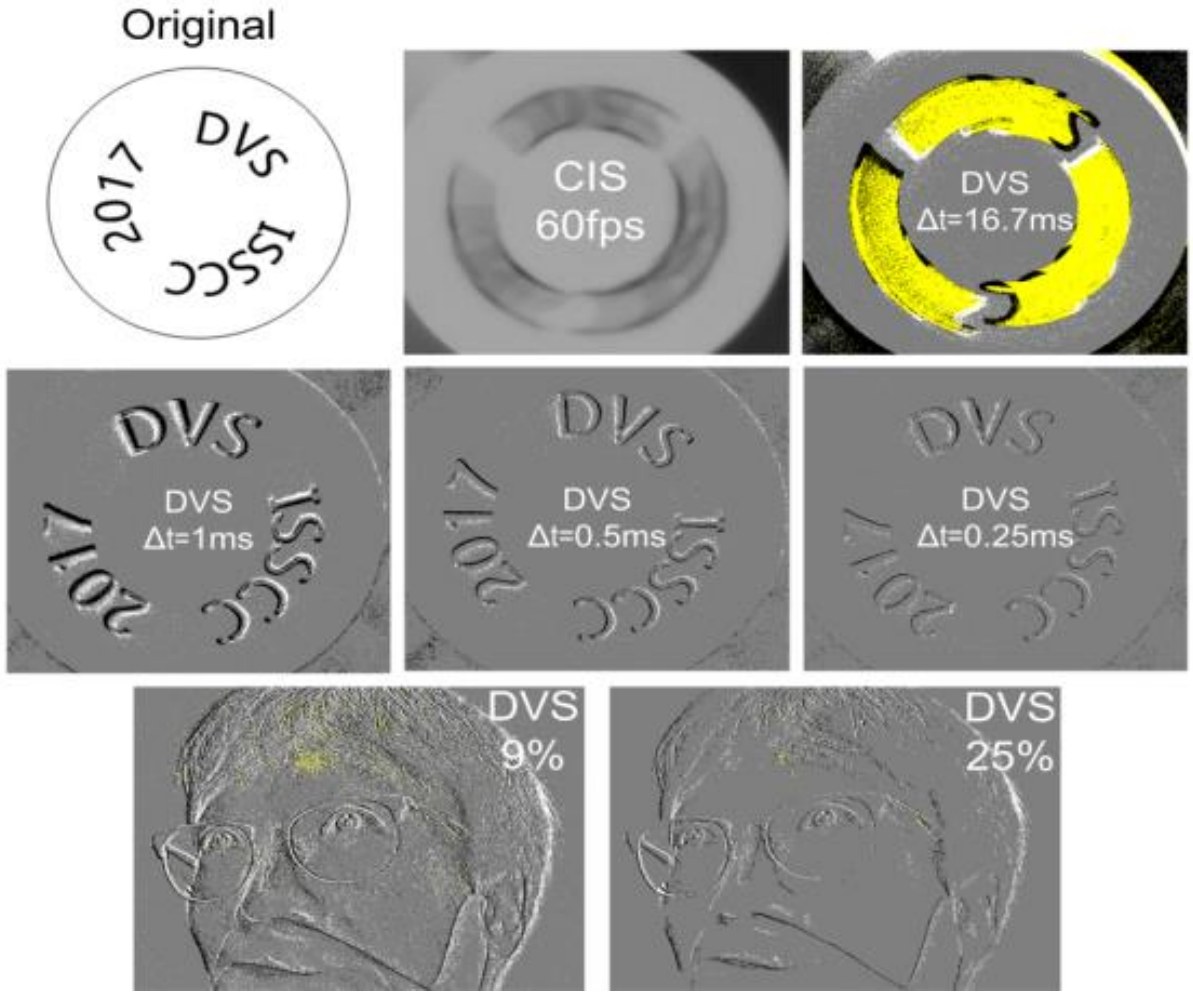
DVS Gen4 (R&D Ver.)



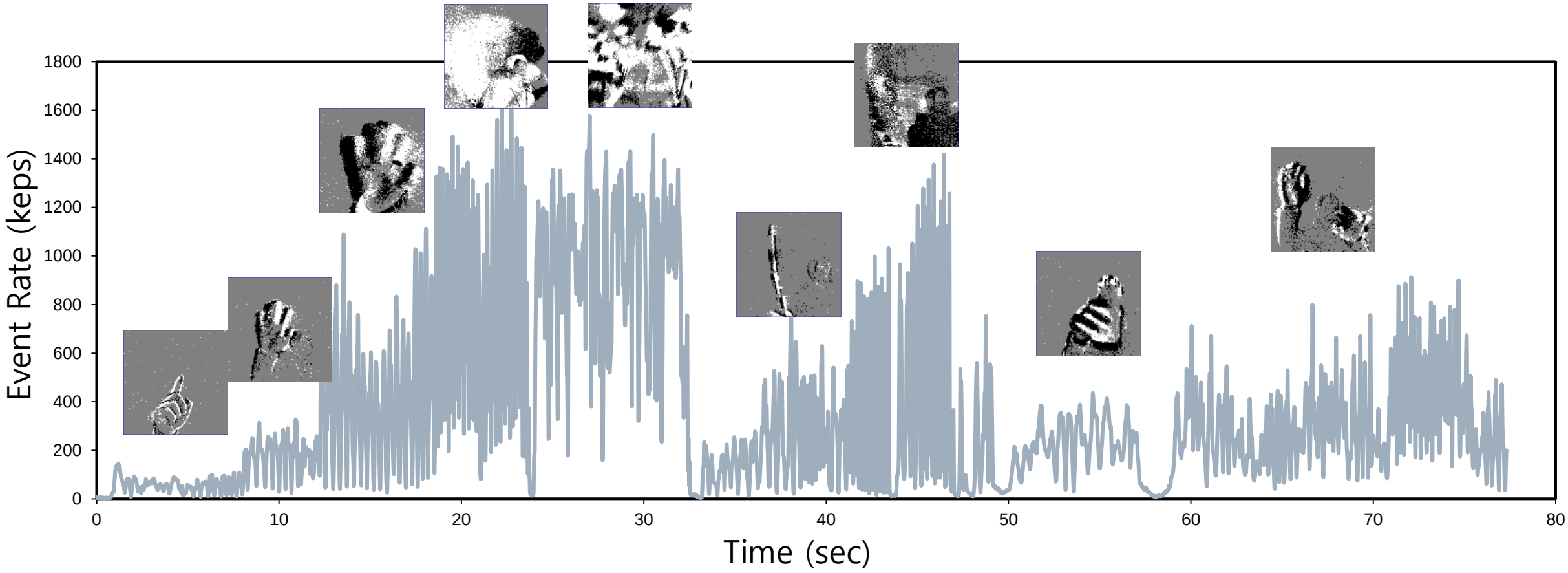
- Brief Specifications
  - 1280 x 960 Pixel Array with 1/2-inch optics
  - Pixel size : 4.95 um
  - Dynamic range : 90 dB (3~100,000 Lux)
  - Frame rate : fixed 1,000 fps (MIPI)
  - Minimum detectable contrast (99.9% response) : < 27.4%
  - Typical power consumption : 140 mW
- Key Features
  - Pixel circuit split using two-stack wafer bonding
  - 2nd CCI supported, CIS RGB Format, De-noise, Anti-Flicker

- Pixel circuit split by using two-stack wafer bonding for smaller pixel size
- Event signal processing block (Anti-flicker, De-noise , etc.) added

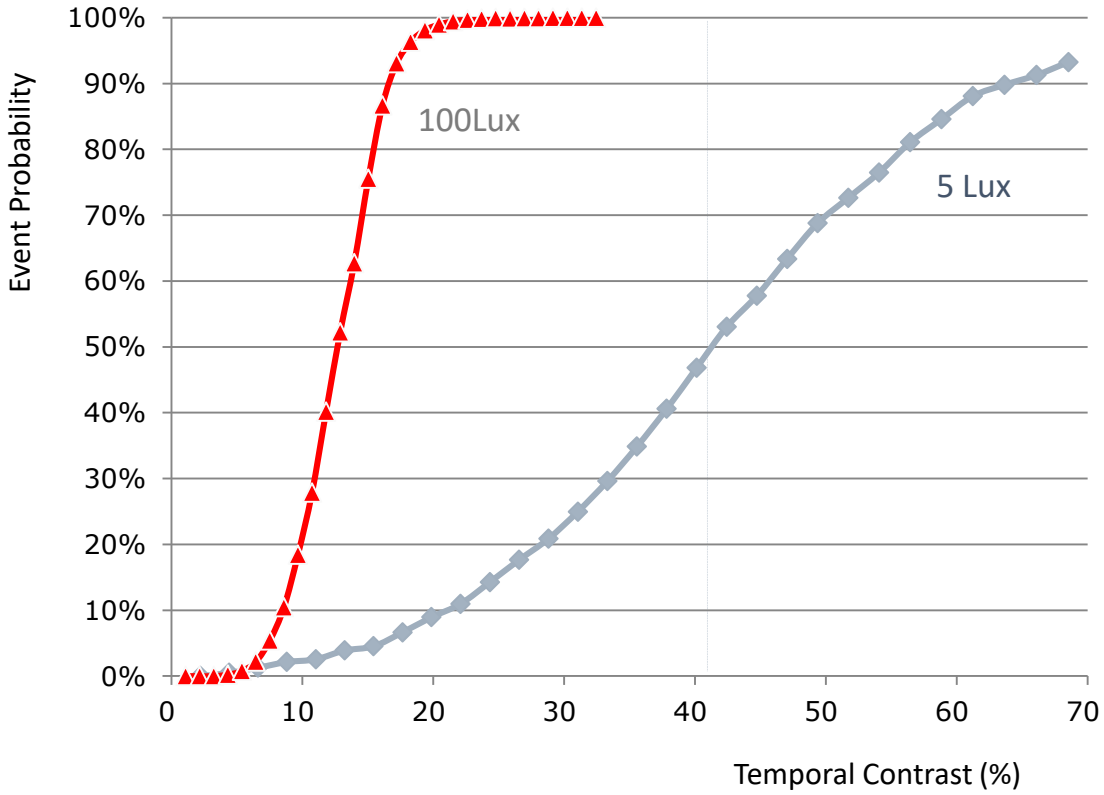
DVS Gen2 Results, only valid for small events

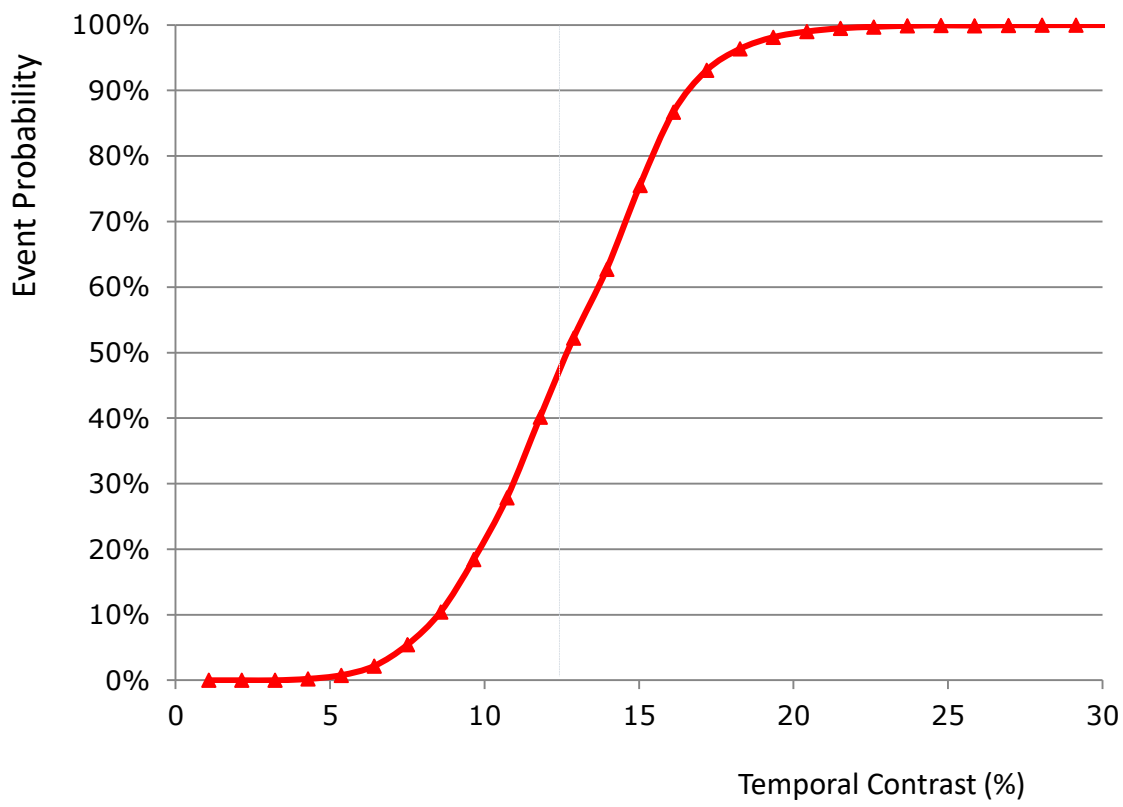


DVS Scene Characteristics

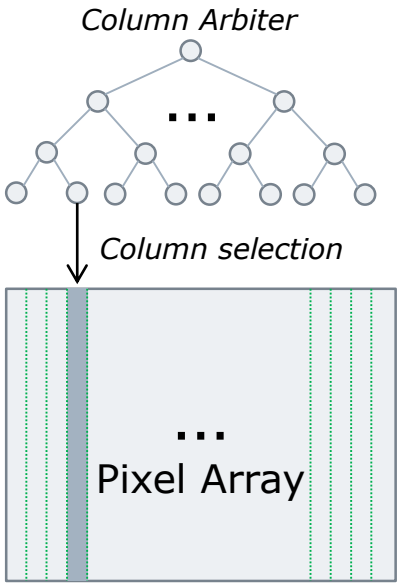


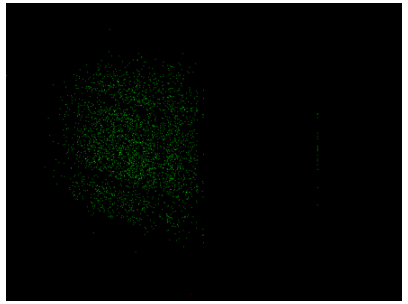
Event Probability

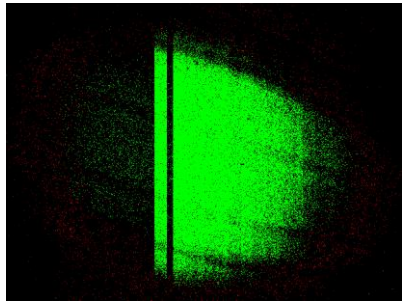


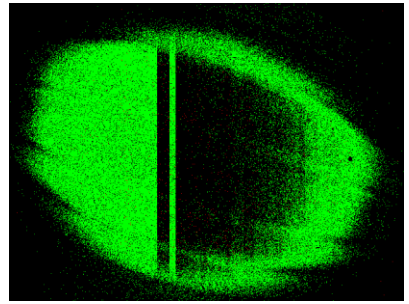


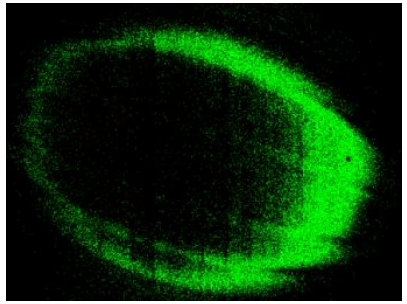
Artifacts and delay by unfair arbitration





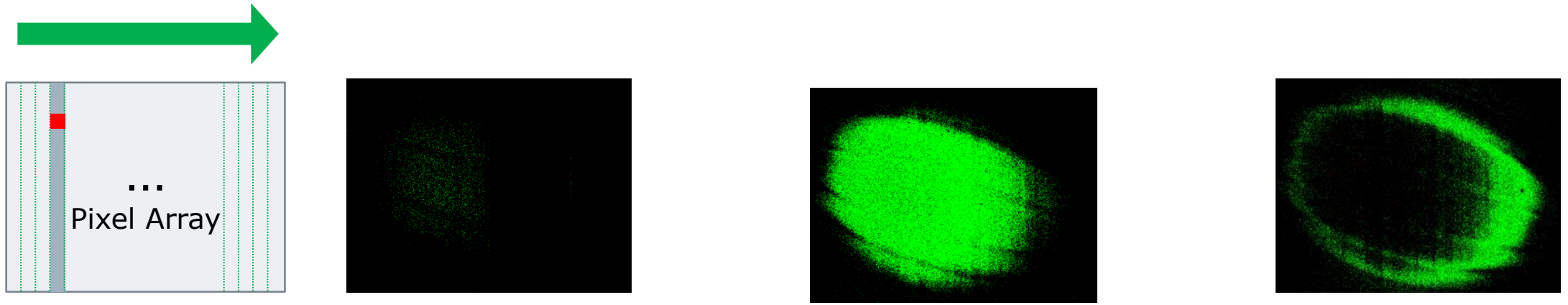






unfair arbitration

Fast column sequential read

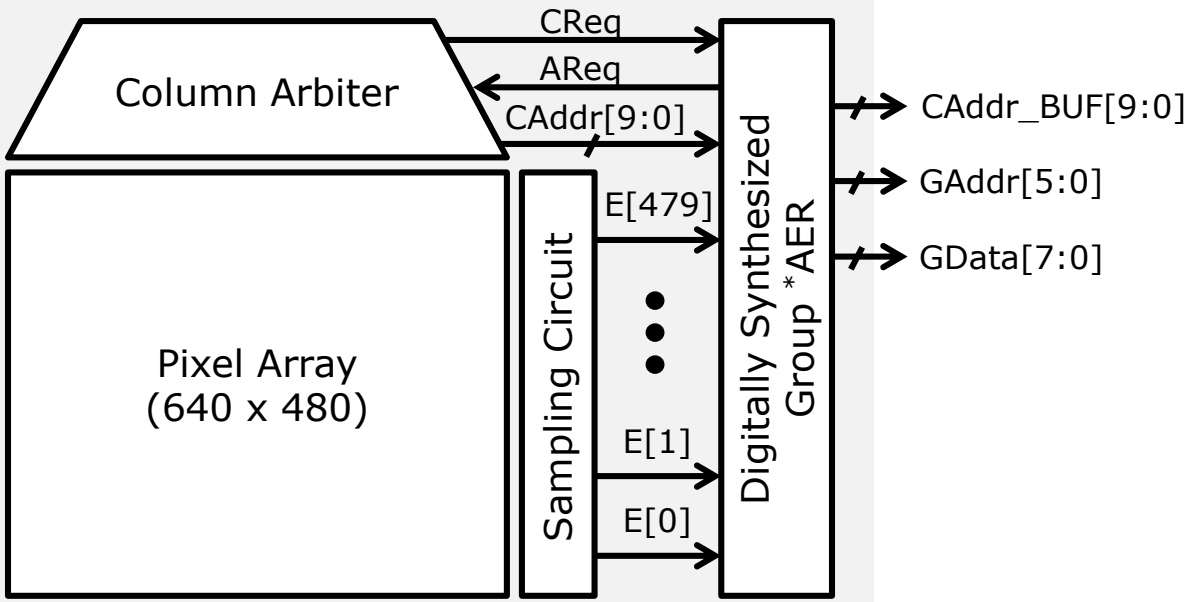


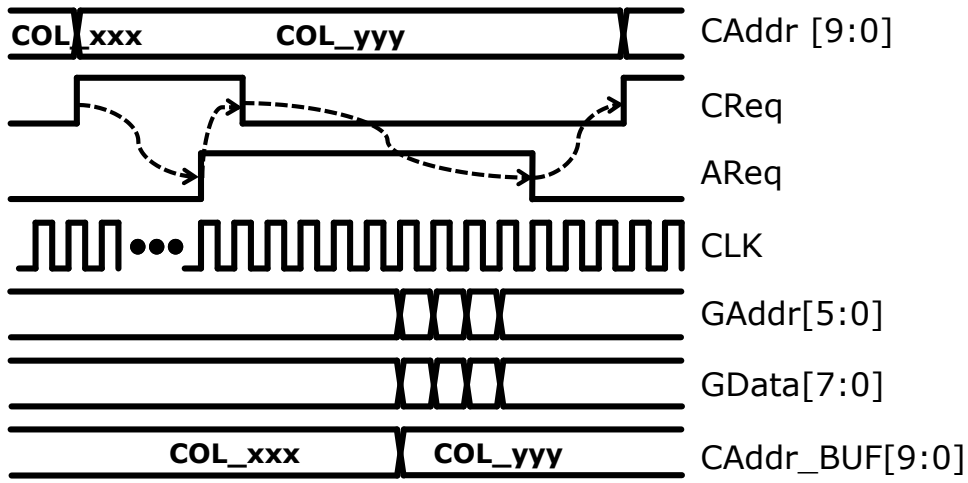
Event scanning
with a capacitor event memory

AER induced latency

- Original AER handles the individual pixel data with address, polarity, and event generation time.
- Group addressing reduces the latency by the interface bandwidth limitation

DVS with Event-Group Handshaking Readout

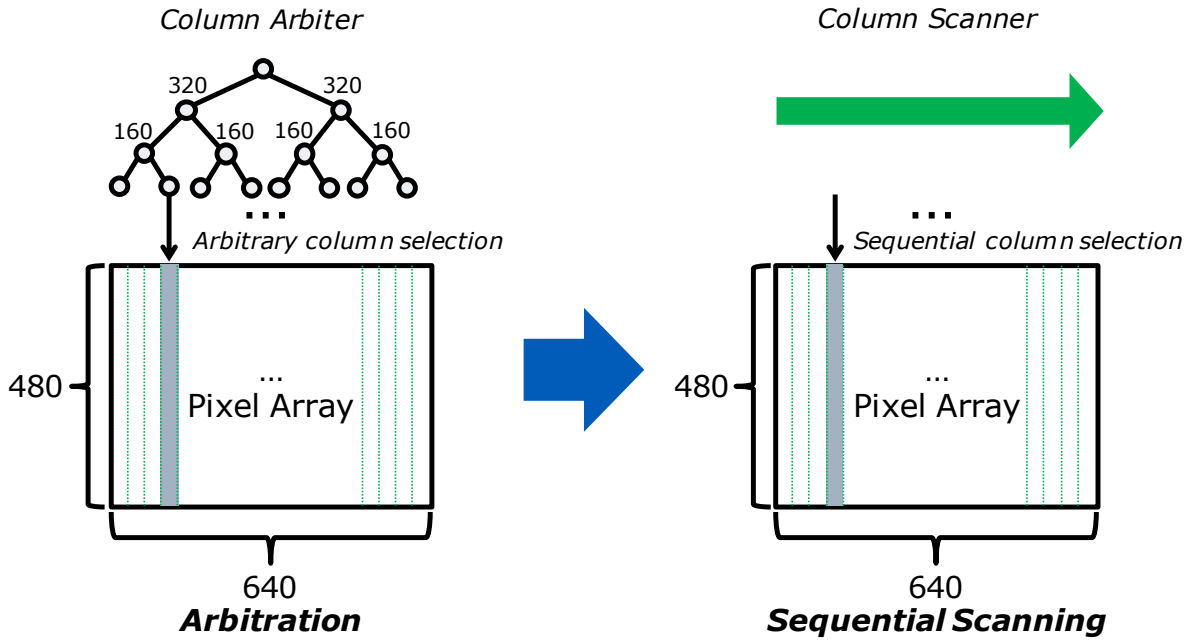


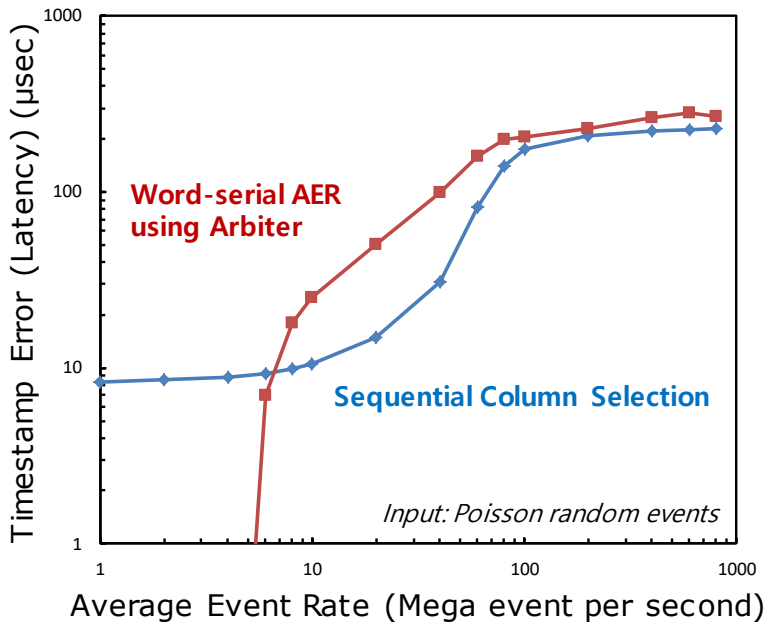


**AER: Address Event Representation*

AER induced motion artifact

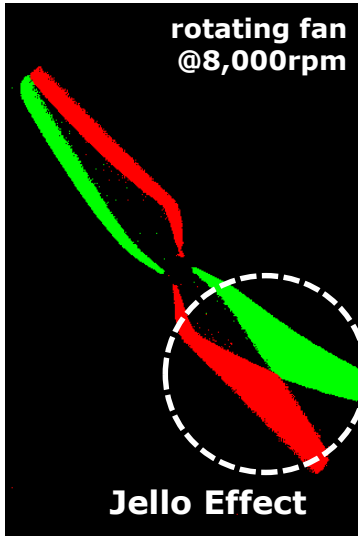
- Image artifact could be induced by the mismatch between event generation time and readout time under high event rate condition



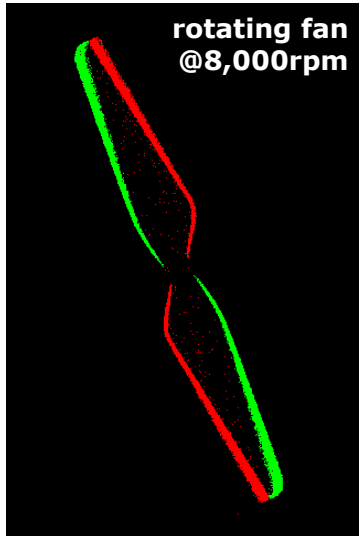


Global Hold, Global Reset

- Global hold is implemented with an event storage in each pixel and its global control signal
- To minimize unwanted tail event, global pixel-reset function is applied



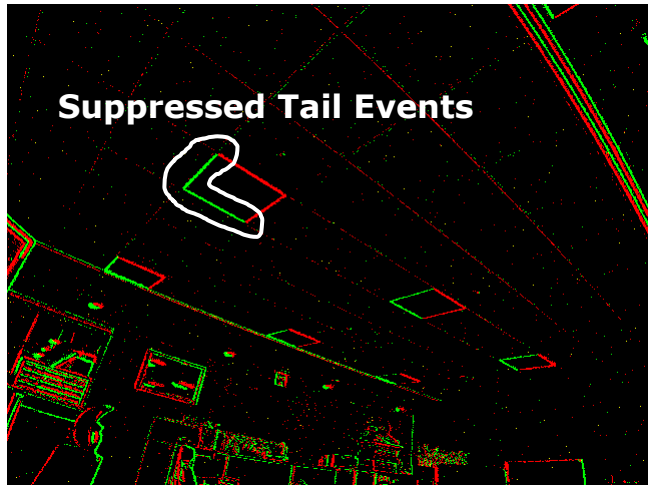
Global Hold Off



Global Hold On

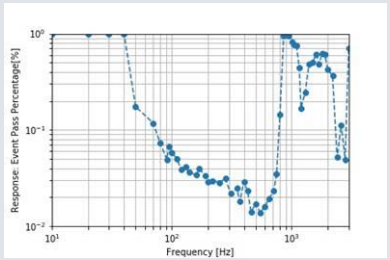
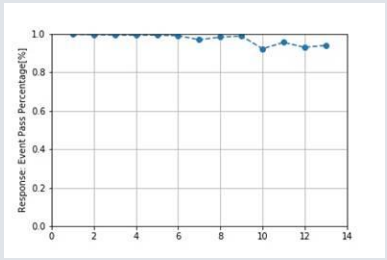
Global Reset Off

One global reset every 32 frames

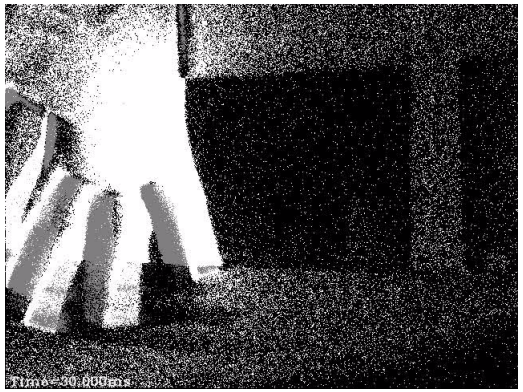


Global Reset On

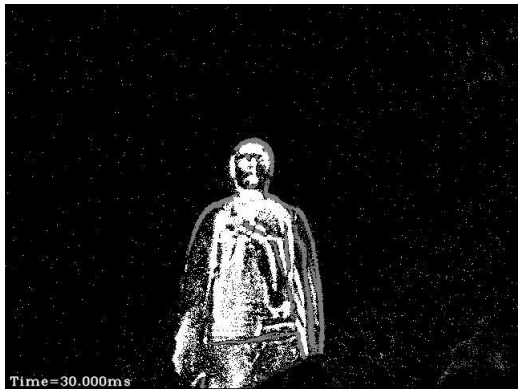
Imbedded Anti-Flicker

Event Pass Rate vs. Flicker Frequency	Event Pass Rate vs. Non-Flicker Recordings	Average Non-flicker Event Pass Percentage
		97.2%

Shaking Hand before Flickering Monitor



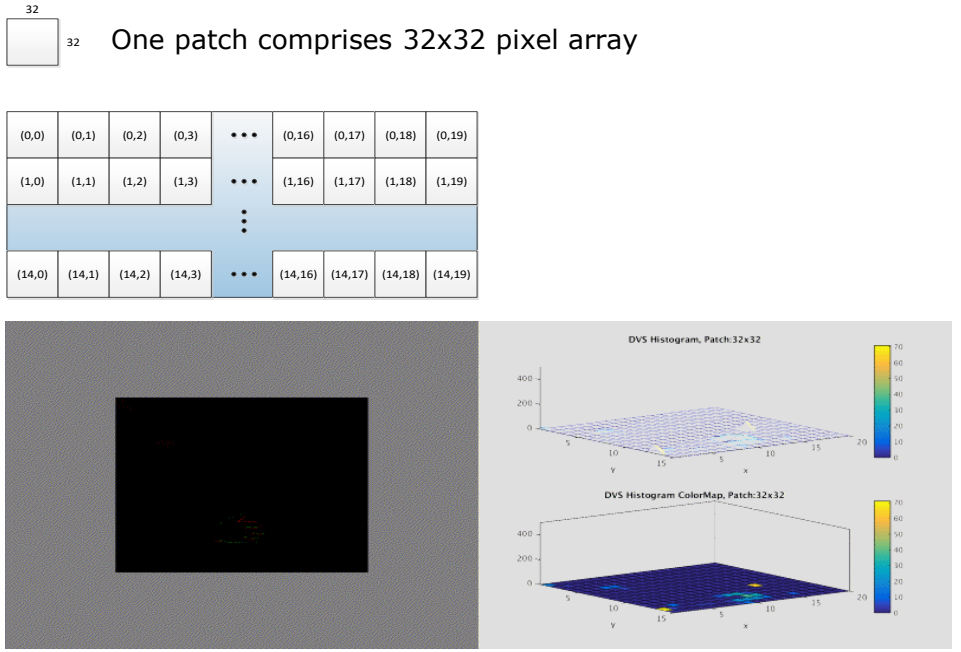
Walking person (AF @non-flickering environment)



Low Power Operation

- Activity Detection

- Spatial Histogram



DVS Human Detection

- **Recognition with securing privacy**

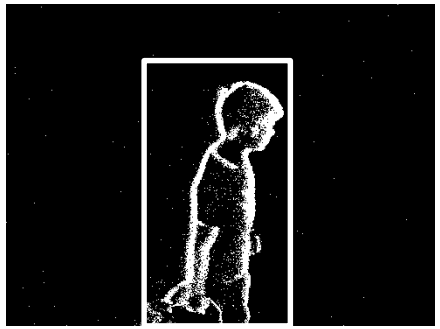
DVS Output
(Edge Image)

Scene

DVS Recognition

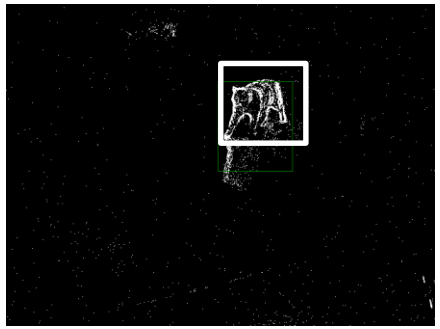
Smaller Database and Faster Training

- 1M-labeled VGA data set, 2.5x faster training and easy converging



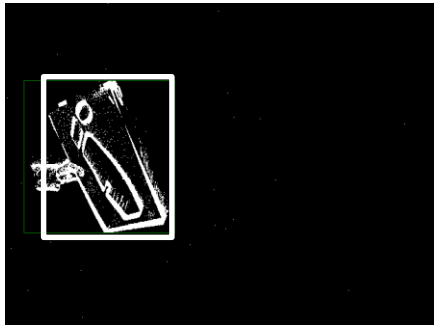
(person, x:230, y:410)

(dog, x:70, y:432)



(cat, x:321, y:125)

(chair, x:322, y:268)



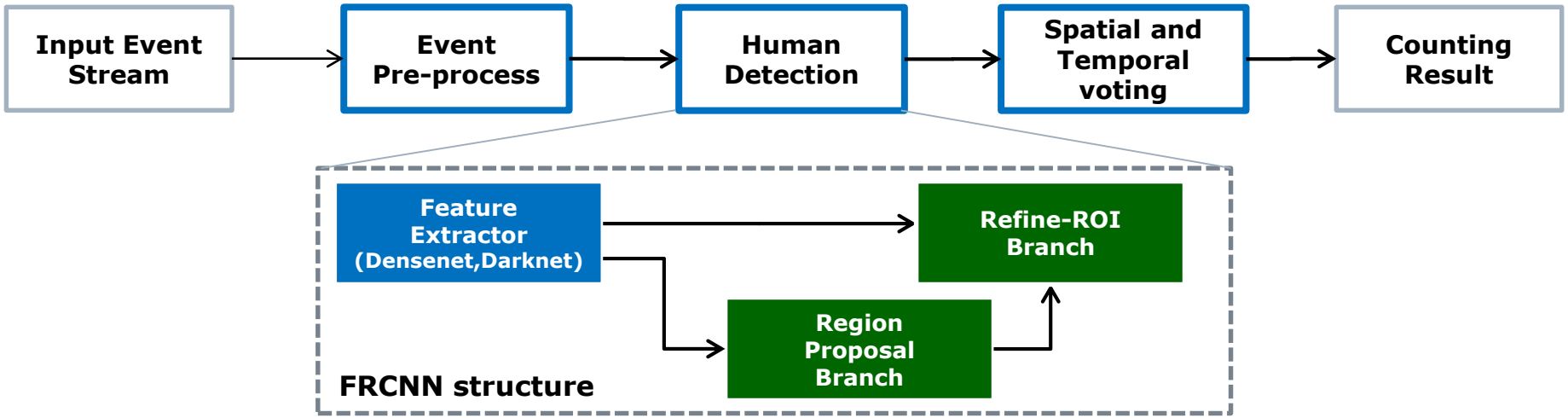
(box, x:120, y:122)

(electric fan, x:621, y:325)

Total 850K images (11 categories) are used for network training.

Algorithm Flow

- C++ based Caffe v1 with CPU only mode (no CUDA lib)

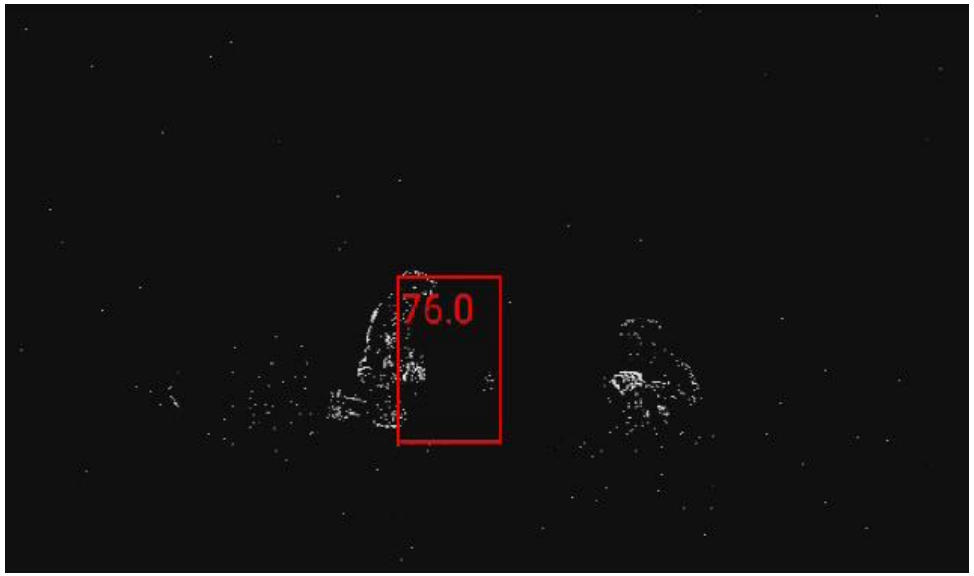


- **DVS event pre-process**
  - Use historic event timestamp of DVS sensor to compensate the sparsity issue.
- **Deep learning Based Human detection**
  - Faster Region Convolutional Neural Network (FRCNN)
  - Calculate the position of objects and classify objects.
- **Spatial and temporal voting**
  - Fusion different detection results to improve accuracy.

Human Detection in Edge Device (ARMv8 Board)

- 92ms @ Exynos 7570, 2 cores (60%)

	Distance	≥ 10 lux	5 ~ 10 lux
Recall	0.5 ~ 1.0m	98%	96%
	1.0 ~ 5.0m	99%	98%
	5.0 ~ 7.0m	96%	92%
False Alarm Rate (FAR)		1%	

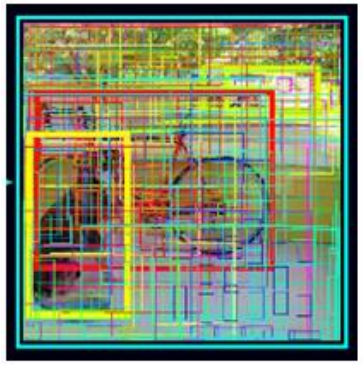


Accuracy Test

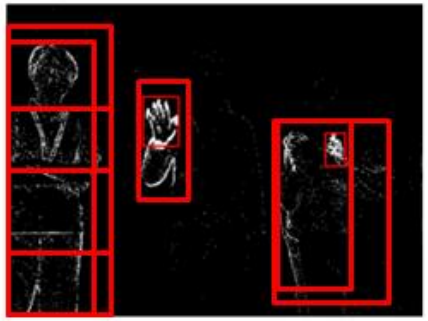
FAR Test

Smaller Network with DVS Images

- **Sparsity** and **binary** features of DVS images
  - Few kernel numbers
  - Few convolutional layers
  - Large stride for layers in the front
- **Small** and **fast** network without accuracy degradation.



CIS



DVS

FRCNN+FPN	# of *Layers	*FLOPs	Processing Time (ms)
CIS-based	91	5.8G	172
DVS-based	24	81M	15

*backbone network

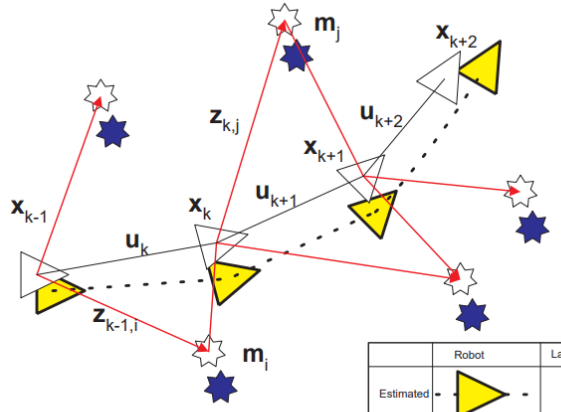
**computed on Titan X

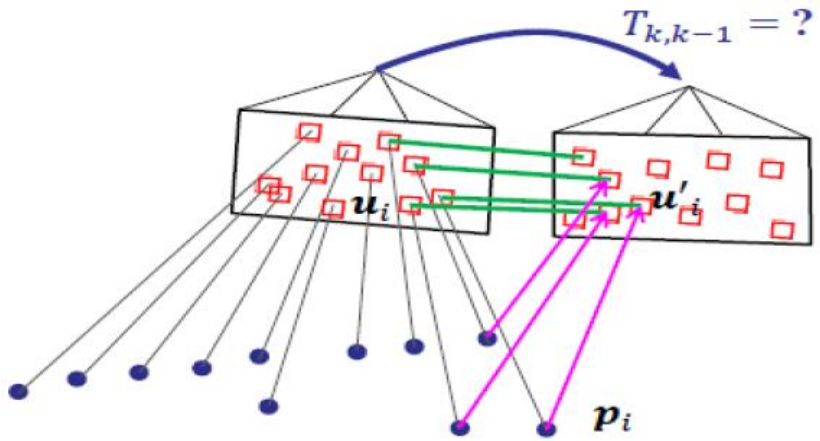
➔ **~11.4x speed up**

- DVS-Based human detection solution (smart home/IoT)
  - Computational speed: 140ms/frame @Exynos 7570(1.4GHz, 8% CPU usages)

DVS SLAM

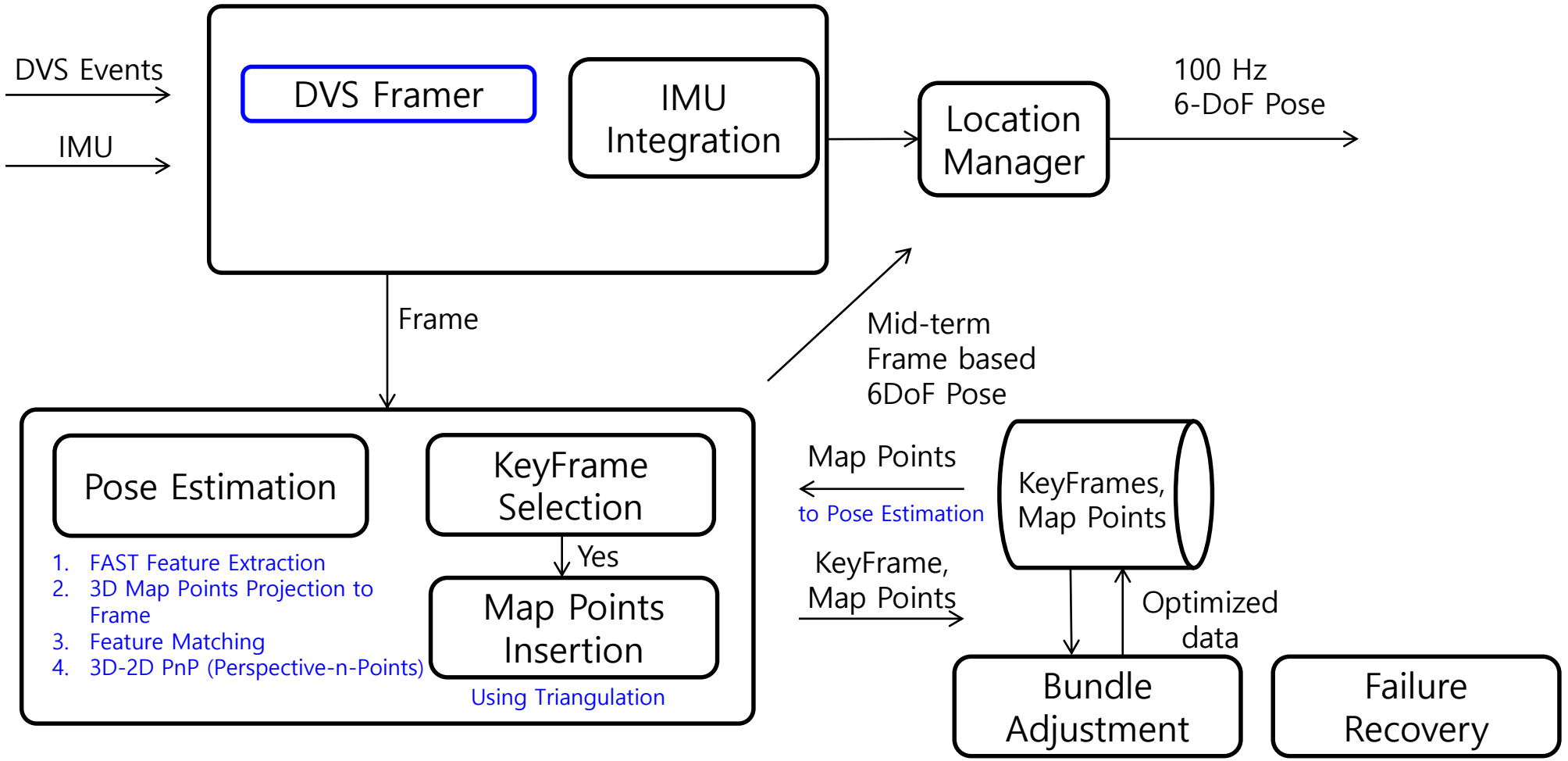
- **VO/VIO/VI-SLAM Algorithms**
  - Track location of device in unknown environment
  - Use visual and inertial sensors
- **Challenges with standard image sensors**
  - Lack robustness to fast motion (blur)
  - Lack robustness to HDR (loss of features)
  - High latency (VR & AR, fast control loops)
- **DVS intrinsic properties deal with these challenges naturally**





Durrant-Whyte, Bailey: "Simultaneous Localisation and Mapping (SLAM) : Part I The Essential Algorithms"
D. Scaramuzza, "Visual Odometry and SLAM: past, present, and the robust-perception age"

DVS-SLAM Algorithm



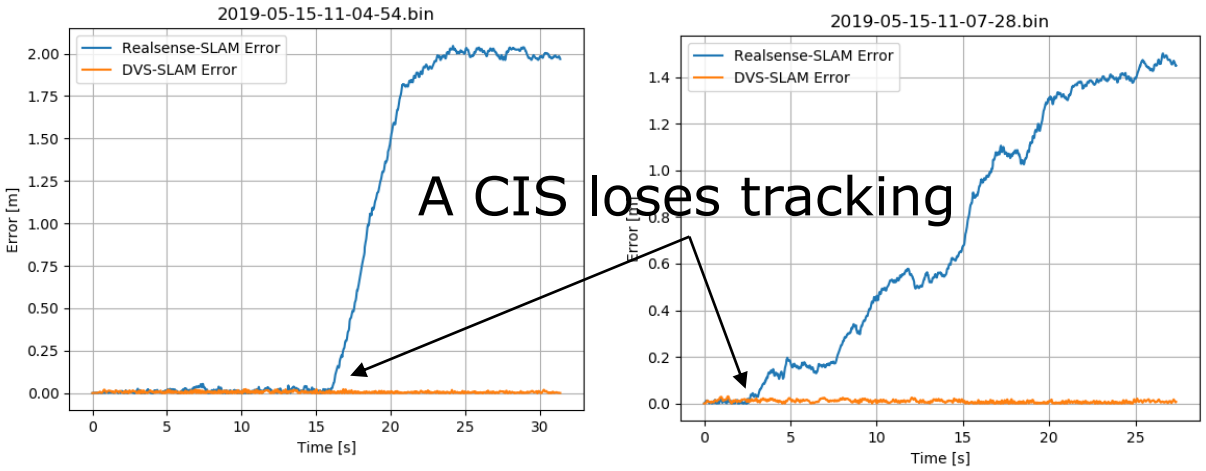
DVS-SLAM Video

DVS-SLAM provides robust tracking even at high-speed motions, thanks to high frame rate (>1000 FPS), sophisticated processing, low computing load with DVS signal characteristics.

Measured Performances

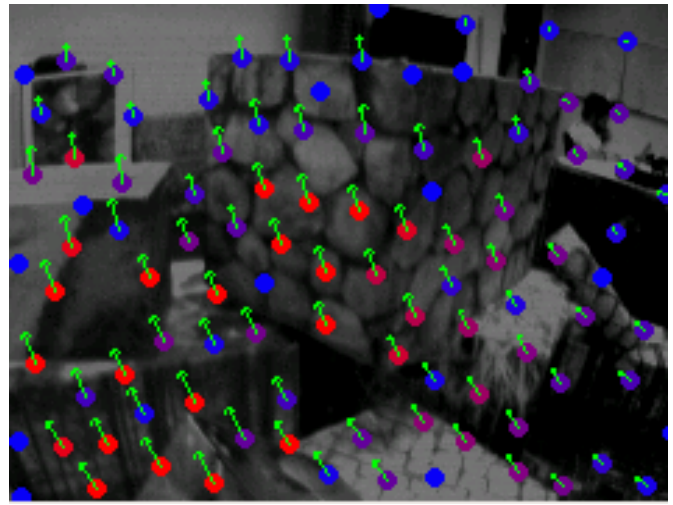
Comparison with a CIS Platform (May 2019)

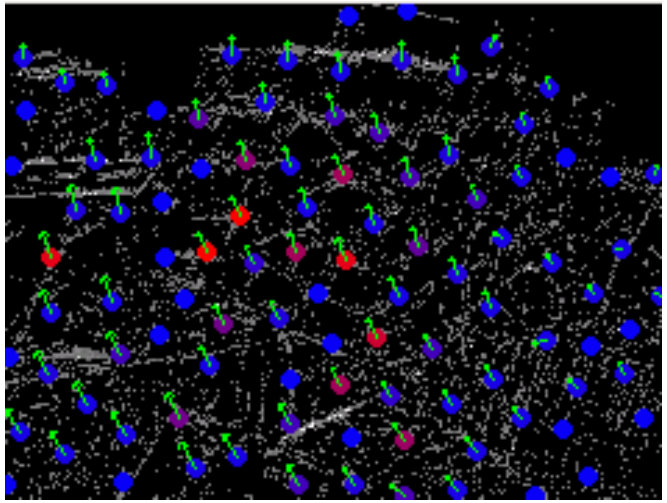
	DVS-SLAM		A CIS Solution	
	Average	Median	Average	Median
Last Position Error (cm)	3.2	2.7	130	125
AAPE (cm)	2.7	2.3	45.4	39.8
Relative Drift Error (%)	0.28%	0.24%	13.61%	12.6%
Relative AAPE (%)	0.25%	0.16%	4.96%	4.02%



DVS+CIS SLAM

- **DVS + CIS: best of both worlds?**
  - DVS provides low latency visual sensing, robustness to fast motion & HDR
  - CIS provides precise long term feature tracking, loop closure & re-localization ability
- **Sample performance of hybrid DVS+CIS SLAM on "The Event-Camera Dataset":**





Scenario	VINS CIS Only AAPE [m]	VINS CIS + DVS AAPE [m]
boxes_6dof	0.29	0.26
boxes_translation	0.09	0.07
hdr_boxes	0.18	0.16
poster_6dof	0.29	0.19
poster_translation	0.02	0.03
dynamic_translation	0.04	0.02
shapes_6dof	0.62	0.49

Summary and discussion

- There were several design changes for the latency, motion artifact, low power operation, smaller pixel size operation, etc.
- DVS based visual recognition shows fewer and faster NN without loss of classification accuracy.
- DVS SLAM shows robustness when the motion is fast, and/or for the high dynamic range scene. DVS with CIS provides better performance than DVS or CIS only cases.
- Remaining issues and research topics
  - bandwidth minimization for higher event rate and spatial resolution
  - improving the pixel sensitivity (response time) at dark illumination condition
  - A way of high level visual information processing combined with SLAM information