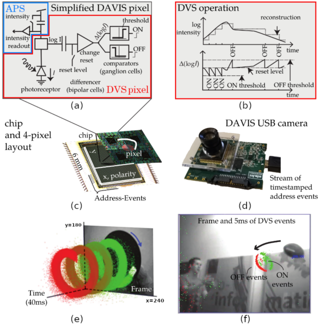
Event-based vision and Deep Learning for dynamic scene analysis

Vincent Frémont

Workshop PPNIV 2022 – Round table

October 23th, 2022, Kyoto, Japan

Event-based camera, Sensing Principle



[Gallego et al., IEEE PAMI 2020]

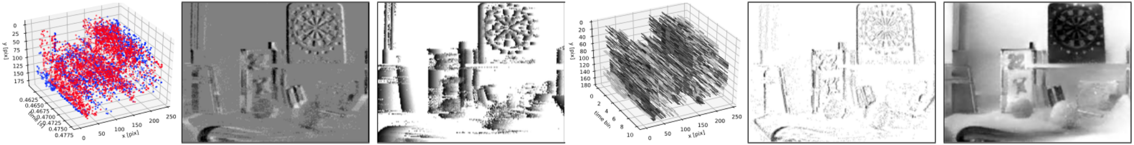
- Response to brightness changes in the scene asynchronously and independently for every pixel.
- Each event **e** is a tuple $< x, y, t, p >$, where (x, y) are the pixel coordinates of events, t is the timestamp of the event, and $p = \pm 1$ is the polarity of the event.
- **Advantages**
  - High Temporal Resolution
  - Low Latency
  - Low Power
  - High Dynamic Range (HDR)
- **Challenges**
  - Different space-time output
  - Different photometric sensing
  - Noise and dynamic effects

Commercial and Prototype Event Cameras

	Supplier	iniVation			Prophesee				Samsung			CelePixel		Insightness
	Camera model	DVS128	DAVIS240	DAVIS346	ATIS	Gen3 CD	Gen3 ATIS	Gen 4 CD	DVS-Gen2	DVS-Gen3	DVS-Gen4	CeleX-IV	CeleX-V	Rino 3
Sensor specifications	Year, Reference	2008 [2]	2014 [4]	2017	2011 [3]	2017 [67]	2017 [67]	2020 [68]	2017 [5]	2018 [69]	2020 [39]	2017 [70]	2019 [71]	2018 [72]
	Resolution (pixels)	128×128	240×180	346×260	304×240	640×480	480×360	1280×720	640×480	640×480	1280×960	768×640	1280×800	320×262
	Latency (μs)	$12\mu s$ @ 1klux	$12\mu s$ @ 1klux	20	3	40 - 200	40 - 200	20 - 150	65 - 410	50	150	10	8	$125\mu s$ @ 10lux
	Dynamic range (dB)	120	120	120	143	> 120	> 120	> 124	90	90	100	90	120	> 100
	Min. contrast sensitivity (%)	17	11	14.3 - 22.5	13	12	12	11	9	15	20	30	10	15
	Power consumption (mW)	23	5 - 14	10 - 170	50 - 175	36 - 95	25 - 87	32 - 84	27 - 50	40	130	-	400	20-70
	Chip size (mm^2)	6.3×6	5×5	8×6	9.9×8.2	9.6×7.2	9.6×7.2	6.22×3.5	8×5.8	8×5.8	8.4×7.6	15.5×15.8	14.3×11.6	5.3×5.3
	Pixel size (μm^2)	40×40	18.5×18.5	18.5×18.5	30×30	15×15	20×20	4.86×4.86	9×9	9×9	4.95×4.95	18×18	9.8×9.8	13×13
	Fill factor (%)	8.1	22	22	20	25	20	> 77	11	12	22	8.5	8	22
	Supply voltage (V)	3.3	1.8 & 3.3	1.8 & 3.3	1.8 & 3.3	1.8	1.8	1.1 & 2.5	1.2 & 2.8	1.2 & 2.8		1.8 & 3.3	1.2 & 2.5	1.8 & 3.3
	Stationary noise (ev/pix/s) at 25C	0.05	0.1	0.1	-	0.1	0.1	0.1	0.03	0.03		0.15	0.2	0.1
	CMOS technology (nm)	350 2P4M	180 1P6M MIM	180 1P6M MIM	180 1P6M	180 1P6M CIS	180 1P6M CIS	90 BI CIS	90 1P5M BSI	90	65/28	180 1P6M CIS	65 CIS	180 1P6M CIS
	Grayscale output	no	yes	yes	yes	no	yes	no	no	no	no	yes	yes	yes
	Grayscale dynamic range (dB)	NA	55	56.7	130	NA	> 100	NA	NA	NA	NA	90	120	50
	Max. frame rate (fps)	NA	35	40	NA	NA	NA	NA	NA	NA	NA	50	100	30
Camera	Max. Bandwidth (Meps)	1	12	12	-	66	66	1066	300	600	1200	200	140	20
	Interface	USB 2	USB 2	USB 3		USB 3	USB 3	USB 3	USB 2	USB 3	USB 3			USB 2
	IMU output	no	1 kHz	1 kHz	no	1 kHz	1 kHz	no	no	1 kHz	no	no	no	1 kHz

[Gallego et al., IEEE PAMI 2020]

Event Representations

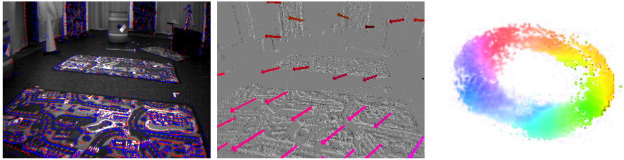


(a) (b) (c) (d) (e) (f)

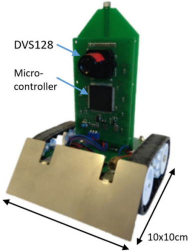
- (a) Events in space time
- (b) Event frame
- (c) Time surface
- (d) Interpolated voxel-grid
- (e) Motion-compensated event image
- (f) Reconstructed intensity image **[Rebecq et al., IEEE PAMI 2019]**

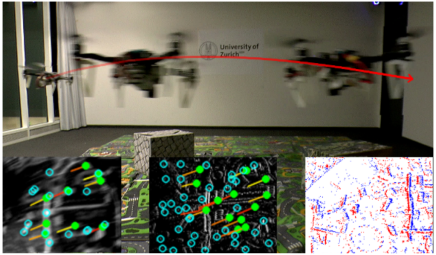
Algorithms

- Feature Detection and Tracking
- Optical Flow Estimation
- 3D reconstruction. Monocular and Stereo
- Pose Estimation and SLAM
- Visual-Inertial Odometry (VIO)
- Image Reconstruction
- Motion Segmentation
- Recognition
- Neuromorphic Control



[Zhu et al., IEEE RAL 2018]



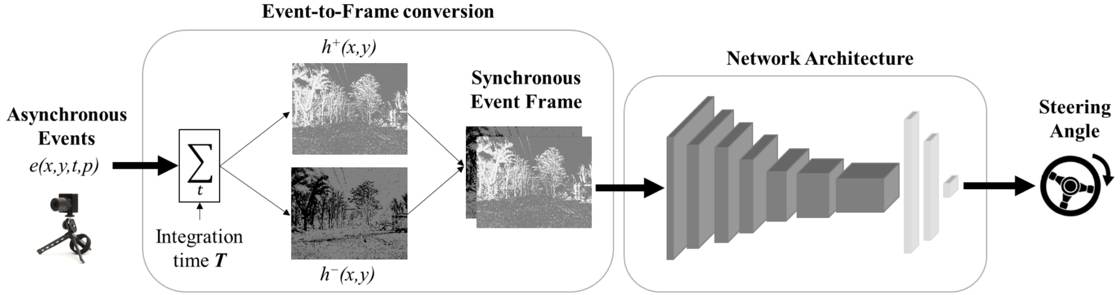


[Rosinol et al., IEEE RAL 2018]

Neuromorphic Control

Event-based Vision meets Deep Learning on Steering Prediction for Self-driving Cars

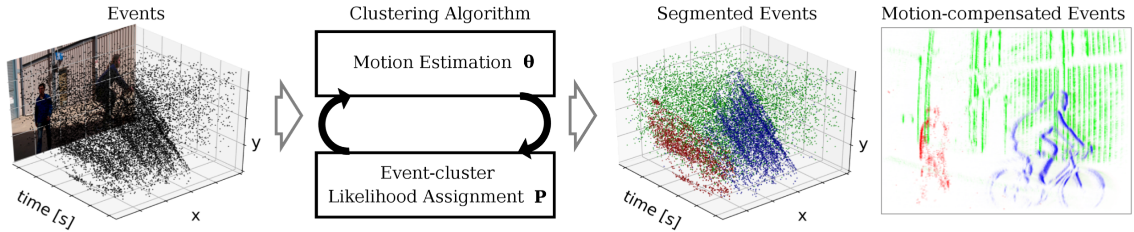
[Maqueda et al., IEEE CVPR 2018]



Motion Compensation

Event-Based Motion Segmentation by Motion Compensation

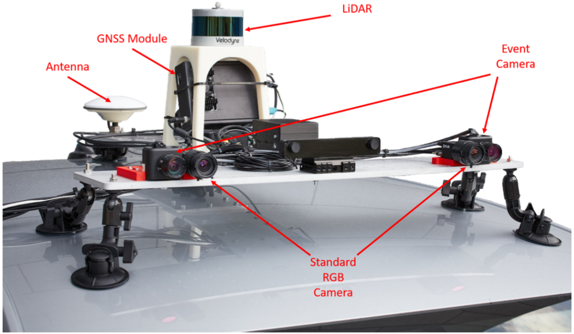
[Stoffregen et al., IEEE ICCV 2019]

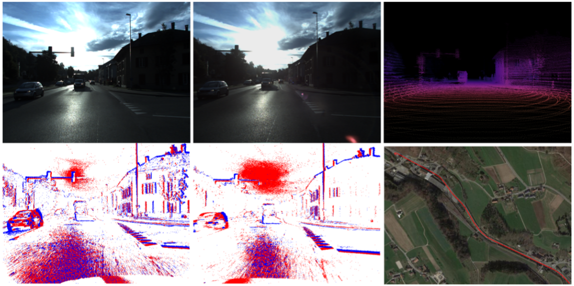


Event Camera Dataset

DSEC: A Stereo Event Camera Dataset for Driving Scenarios

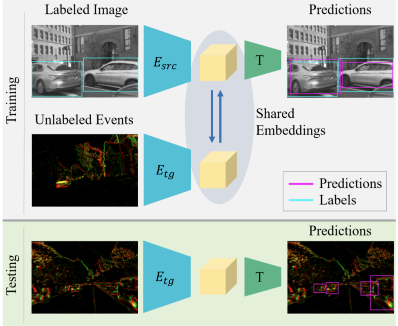
[Gehrig et al., IEEE RAL 2021]

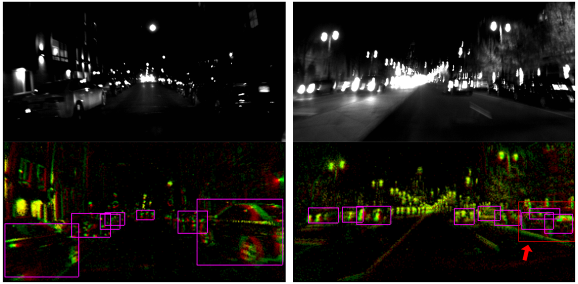




Object Detection

Unsupervised Domain Adaptation

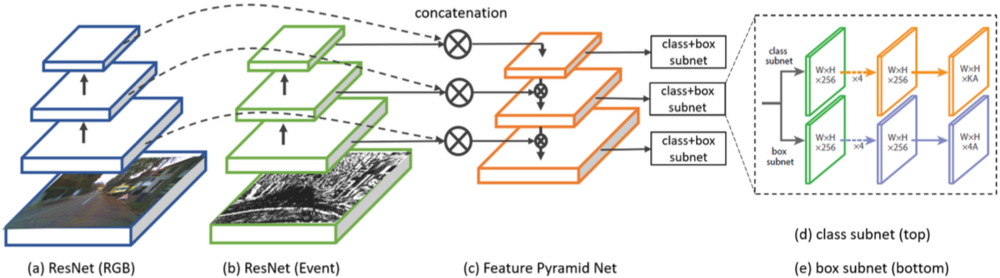




[Messikommer et al., IEEE RAL 2021]

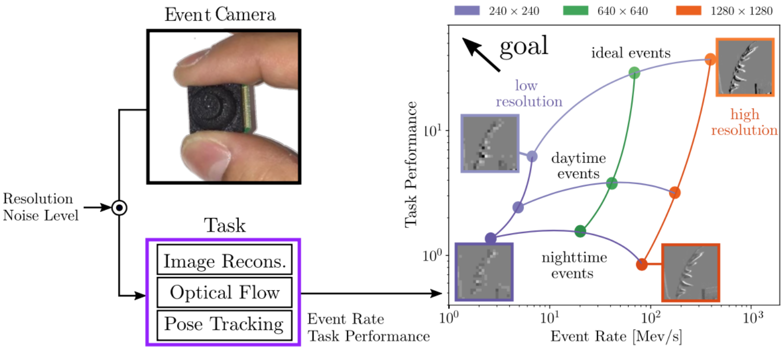
Object Detection

Object Detection in Adverse Conditions



[Tomy et al., IEEE ICRA 2022]

Impact of Camera Resolution

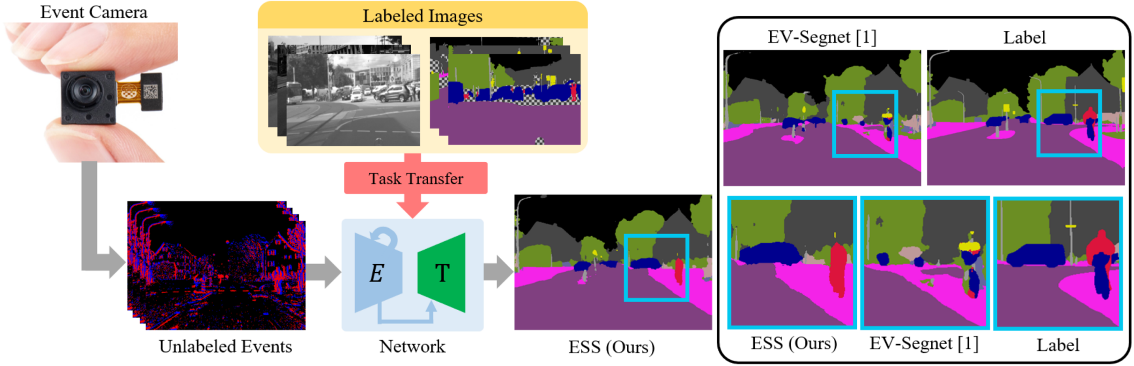


[Gehrig et al., arXiv 2022]

- Higher resolution sensors significantly increase the required data bandwidth, while burdening downstream tasks
- Higher resolution cameras are also more sensitive to temporal effects such as slow pixel response times
- When used in challenging high-speed scenarios and in low light, lower- resolution sensors often show a better performance while using lower bandwidth

Semantic Segmentation

Unsupervised Domain Adaptation



[Sun et al., ECCV 2022]

Conclusion

- Event cameras pose a paradigm shift in the way visual information is acquired, so many challenges arise
- Many advantages over frame-based cameras: low latency, low power, high speed and high dynamic range
- Event cameras pose the challenge of rethinking perception, decision and control, especially when using deep learning approaches
- In autonomous vehicles applications, event-based vision will be helpful to catch the dynamic nature of the scene and to increase the robustness against adverse conditions
- Open questions: Event representations, end-to-end Deep Neural Networks architectures, training data, dedicated hardware processing, multi-sensors architecture, etc.