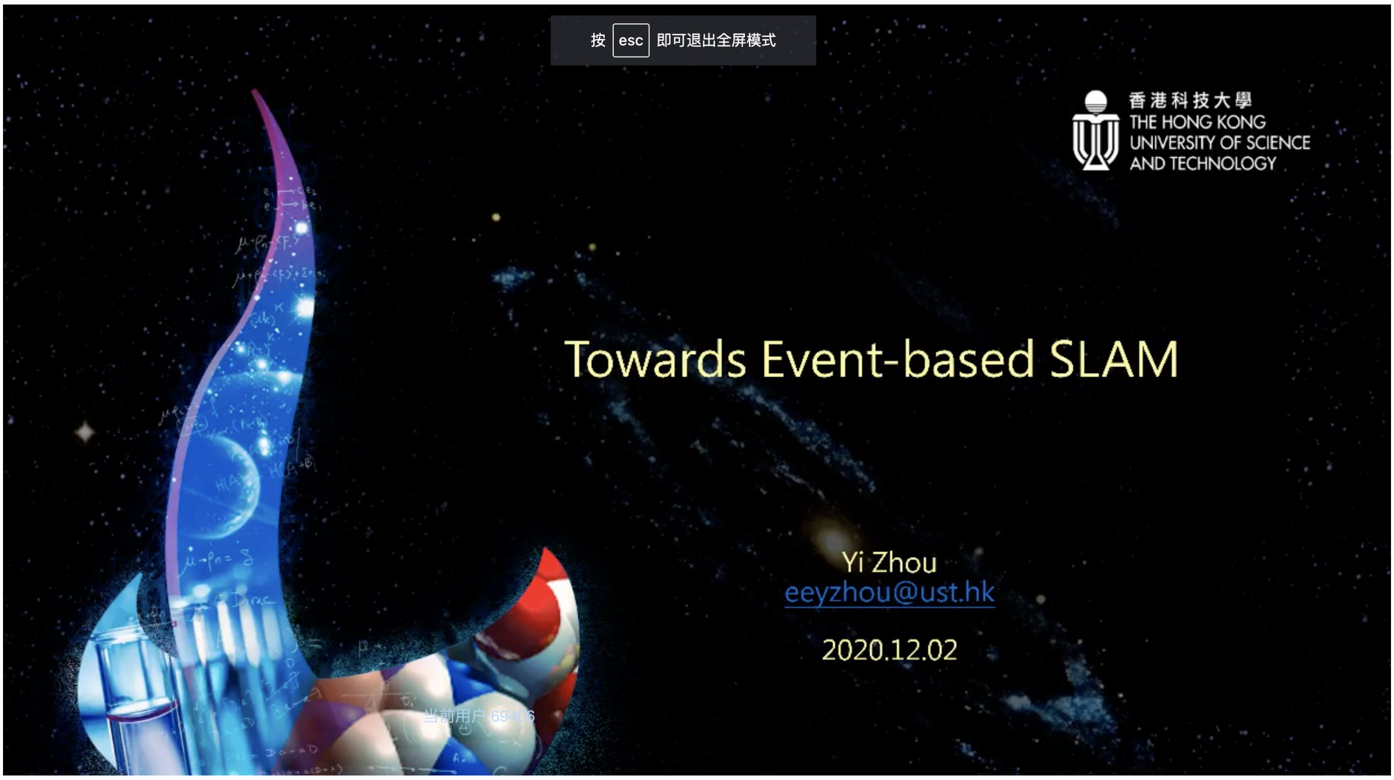
Towards Event-based SLAM



Towards Event-based SLAM

Outline

- Introduction to Event-based Cameras
- Event-based SLAM/VO
- ESVO System
- Conclusion

DVS 128, by Zurich-based IniLabs Ltd.

Open Challenges for Standard Cameras

Latency & Motion Blur

Dynamic Range

Image courtesy: Tutorial on Event-based Camera by D. Scaramuzza

Open Challenges for Standard Cameras

Latency & Motion Blur

Dynamic Range

Rescue Mission

Image courtesy: Tutorial on Event-based Camera by D. Scaramuzza

Open Challenges for Standard Cameras

Latency & Motion Blur

Dynamic Range

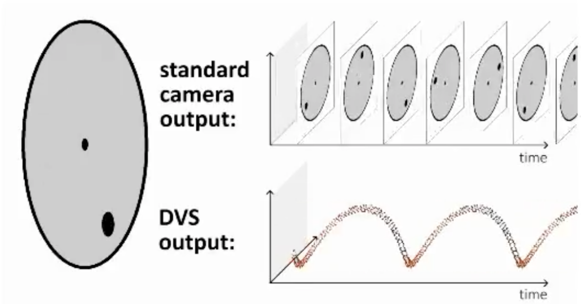
Rescue Mission

Event cameras do not suffer from these problems!

Image courtesy: Tutorial on Event-based Camera by D. Scaramuzza

Properties and Working Principle

- ◆Novel sensor that measures only **motion in the scene**
- ◆**Low-latency** (~ 1 μs)
- ◆**No motion blur**
- ◆**Asynchronous and independent pixels**
- ◆**High dynamic range** (140 dB instead of 60 dB)
- ◆**Ultra-low power** (mean: 1mW vs 1W)

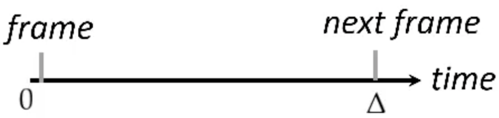


Video from here: https://youtu.be/LauQ6LWTkxM?t=30

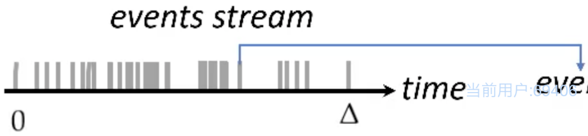
Lichtsteiner, Posch, Delbruck, A 128x128 120 dB 15μs Latency Asynchronous Temporal Contrast Vision Sensor, 2008

Standard Camera v.s. Event Camera

- A **traditional camera** outputs frames at **fixed time intervals**:



- By contrast, a **DVS** outputs **asynchronous events** at ***microsecond* resolution**. An event is generated each time a single pixel detects an intensity changes value

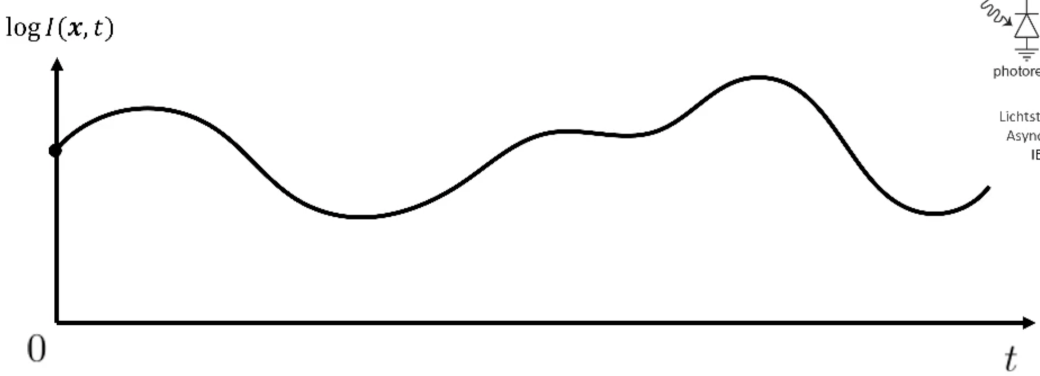


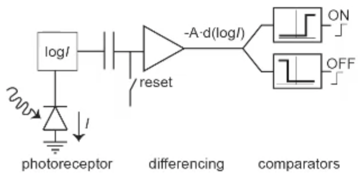
event: $\left\langle t, \langle u,v\rangle, sign\left(\frac{dI(u,v)}{dt}\right)\right\rangle$

Lichtsteiner, Posch, Delbruck, A 128x128 120 dB 15μs Latency Asynchronous Temporal Contrast Vision Sensor, 2008

Generative Event Model

Consider the intensity at a **single pixel…**





Lichtsteiner et al., A 128x128 120 dB 15μs Latency Asynchronous Temporal Contrast Vision Sensor, IEEE Journal of Solid-State Circuits, 2008

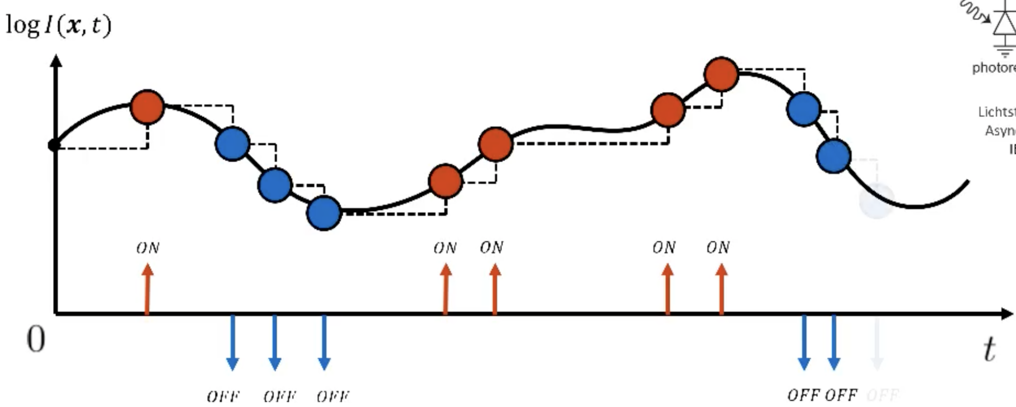
Events are triggered **asynchronously**

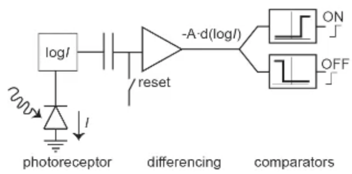
Image courtesy: Tutorial on Event-based Camera by D. Scaramuzza

Generative Event Model

Consider the intensity at a **single pixel…**

$$\pm C = \log I(\boldsymbol{x}, t) - \log I(\boldsymbol{x}, t - \Delta t)$$





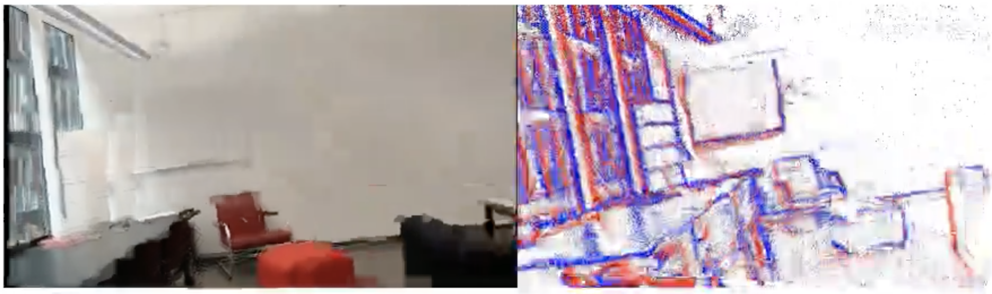
Lichtsteiner et al., A 128x128 120 dB 15μs Latency Asynchronous Temporal Contrast Vision Sensor, IEEE Journal of Solid-State Circuits, 2008

Events are triggered **asynchronously**

Image courtesy: Tutorial on Event-based Camera by D. Scaramuzza

Event Camera Output

Standard Camera — Event Camera (ON, OFF events)



ΔT = 40 ms

Applciations

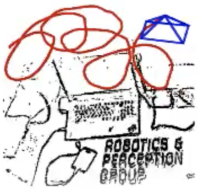
- **Internet of Things (IoT)**
  - Low-power, always-on devices for monitoring and surveillance
- **Automotive**:
  - low-latency, high dynamic range (HDR) object detection
  - low-power training & inference
  - low-memory storage
- **AR/VR**
  - low-latency, low-power tracking
- **Industrial automation**
  - Fast pick and place

Information source: Tutorial on Event-based Camera by D. Scaramuzza

Products and Suppliers

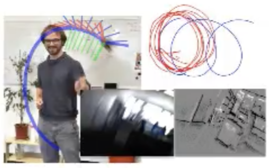
	Sensors	Resolution	Price
Inivation	Frames, events, IMU	QVGA (346 x 260)	6,000 USD
Insightness	Frames, events, IMU	QVGA (346 x 262)	6,000 USD
Prophesee	Events, IMU, absolute intensity	1M pixels	4,000 USD
CelexPixel Technology	Events, IMU, absolute intensity	1M pixels	1,000 USD
Samsung Electronics	Events, IMU	Up to 1M pixel	Not listed

Event-based SLAM/VO

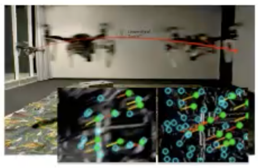


EVO [RAL 17]

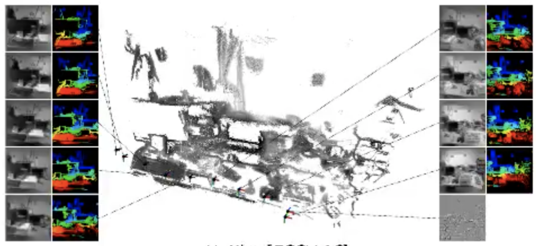
G. Gallego *et. al* [T-PAMI 18]



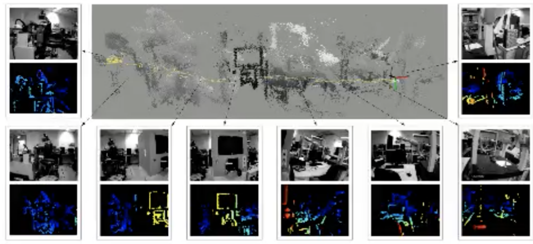
H. Rebecq [*BMVC 17*]



Ultimate SLAM [*RAL 18*]

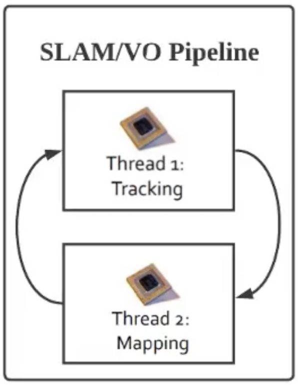


H. Kim [*ECCV 16*]



ESVO [*Arxiv 20*]

Review on Event-based Methods



PTAM *[ISMAR 07]*

- ☐ Event-based Depth Estimation (3D Reconstruction)

  [*ISVC 11, TNN 12, Front. Neurosci. 14, 18, Meas. Sci. Technol. 14, Neural Proc. Lett. 16, Sci. Rep. 17, Front. Neurorobot. 19 , IJCV 18*]

- ☐ Event-based Camera Pose Estimation

  [*RSS 15, TPAMI 18, RAL 17, ICRA 19 , IJCNN 11, BMVC 14, ICCP 17, ROBIO 12, ICVS 13, IROS 14*]

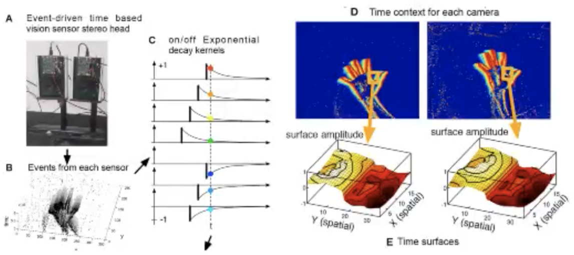
- ☐ Event-based VO/SLAM Systems

  [*ECCV 16, RAL 17*]

Event-based Depth Estimation (3D Reconstruction)

☐ Instantaneous Stereo

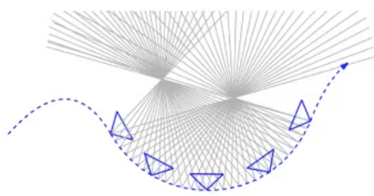
Two-Step paradigm
① Finding Eipolar matching
② Triangulation



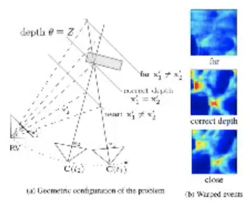
SH. Ieng, *et. al.*, Neuromorphic Event-Based Generalized Time-Based Stereovision, Front. Neurosci. 2018

☐ Temporal Stereo (monocular event camera!)

① Require knowledge of the camera motion
② Integrate information from the events over a longer time interval



[*IJCV 18*]



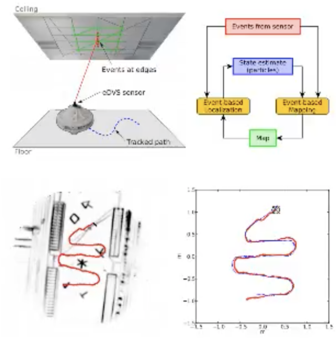
[*CVPR 18*]

H. Rebecq, et. al., "EMVS: Event-based multi-view stereo—3D reconstruction with an event camera in real-time," IJCV. 2018.
G. Gallego, *et, al.*, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation," CVPR 2018

Event-based Camera Pose Estimation

Motion complexity : simple -> complex

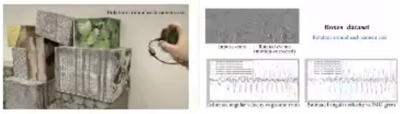
Planar Motion



D. Weikersdorfer, et.al, "Simultaneous localization and mapping for event-based vision systems," ICVS, 2013.

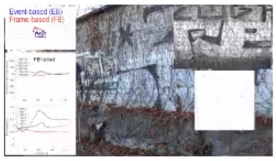
3D Rotation

H. Kim, et. al., "Simultaneous mosaicing and tracking with an event camera," BMVC, 2014



G. Gallego, et. al., "Accurate angular velocity estimation with an event camera," RAL 2017

6-DoF Motion



G. Gallego, et. al., "Event-based, 6-DOF camera tracking from photometric depth maps," T-PAMI 2018.

S. Bryner, et. al, "Event-based, direct camera tracking from a photometric 3D map using nonlinear optimization," ICRA 2019

Event-based SLAM/VO Systems

Early Works (Quick-Hand Solutions)

Censi, A., Scaramuzza, D.: Low-Latency Event-Based Visual Odometry. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) 2014. (CMOS + Event Camera)

Weikersdorfer, D., Adrian, D.B., Cremers, D., Conradt, J.: Event-based 3D SLAM with a depth-augmented dynamic vision sensor. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) 2014. (RGB-D + Event Camera)

Event-based SLAM/VO Systems

H. Kim, S. Leutenegger, and A. J. Davison, "Real-time 3D reconstruction and 6-DoF tracking with an event camera," in Eur. Conf. Comput. Vis. (ECCV), 2016. Best Paper Award!

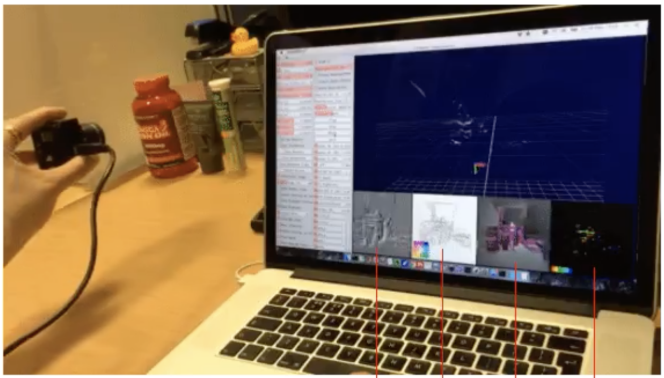
H. Rebecq, T. Horstschafer, G. Gallego, and D. Scaramuzza, "EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real-time," IEEE Robot. Autom. Lett. (RAL), 2017.

Real-time 3D reconstruction and 6-DoF tracking with an event camera [ECCV 16]

Method Outline

Three interleaved probabilistic filters (EKFs)

- Filter 1: Tracks global 6-DoF camera motion
- Filter 2: Estimates the log intensity gradients in a keyframe image
- Filter 3: Estimates the inverse depths of a keyframe



Video courtesy:
https://www.youtube.com/watch?v=yHLyhdMSw7w&ab_channel=HanmeKim

Real-time 3D reconstruction and 6-DoF tracking with an event camera [ECCV 16]

Filter1: Tracks global 6-DoF camera motion

Predict

$$
\mathbf{x}^{(t|t-\tau)} = \mathbf{x}^{(t-\tau|t-\tau)} + \mathbf{n} \,, \qquad \mathrm{P}_{\mathbf{x}}^{(t|t-\tau)} = \mathrm{P}_{\mathbf{x}}^{(t-\tau|t-\tau)} + \mathrm{P}_{\mathbf{n}} \,,
$$

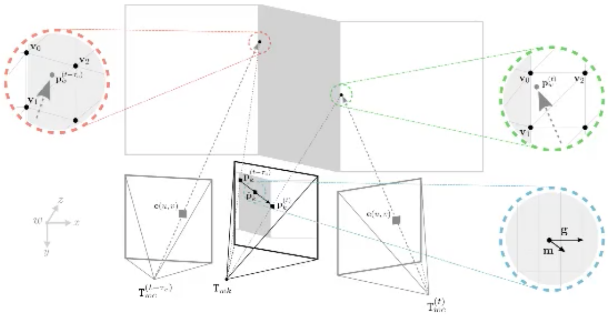
Constant position motion model

Measurement Model

$$
z_{\mathbf{x}} = \pm C \,, \quad h_{\mathbf{x}}(\mathbf{x}^{(t|t-\tau)}) = \mathrm{I}_l\left(\mathbf{p}_w^{(t)}\right) - \mathrm{I}_l\left(\mathbf{p}_w^{(t-\tau_c)}\right) \,,
$$
$$
\text{where } \mathrm{I}_l(\mathbf{p}_w) = (1-a-b)\mathrm{I}_l(\mathbf{v}_0) + a\mathrm{I}_l(\mathbf{v}_1) + b\mathrm{I}_l(\mathbf{v}_2) \,.
$$

Update

$$
\mathbf{x}^{(t|t)} = \mathbf{x}^{(t|t-\tau)} + \mathbf{W}_{\mathbf{x}}\nu_{\mathbf{x}} \,, \qquad \mathrm{P}_{\mathbf{x}}^{(t|t)} = \left(\mathrm{I}_{6\times 6} - \mathbf{W}_{\mathbf{x}}\frac{\partial h_{\mathbf{x}}}{\partial \mathbf{x}^{(t|t-\tau)}}\right)\mathrm{P}_{\mathbf{x}}^{(t|t-\tau)}
$$



Basic geometry for tracking, inverse depth estimation and gradient estimation.

H. Kim, et. al, ECCV' 2016

Real-time 3D reconstruction and 6-DoF tracking with an event camera [ECCV 16]

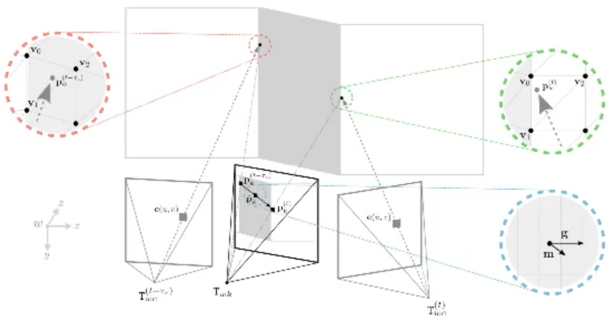
Filter2: Pixel-Wise EKF Based Gradient Estimation

Measurement Model

$$
z_{\mathbf{g}} = \pm\frac{C}{\tau_c} \,, \quad h_{\mathbf{g}} = (\mathbf{g}(\hat{\mathbf{p}}_k)\cdot\mathbf{m}) \,, \quad \text{where } \mathbf{m} = \frac{\mathbf{p}_k^{(t)} - \mathbf{p}_k^{(t-\tau_c)}}{\tau_c}
$$

Log Intensity Reconstruction

$$
\min_{\mathrm{I}_l}\left\{\int_\Omega ||\mathbf{g}(\mathbf{p}_k) - \nabla\mathrm{I}_l(\mathbf{p}_k)||^h_{\epsilon_d} + \lambda||\nabla\mathrm{I}_l(\mathbf{p}_k)||^h_{\epsilon_r}\,\mathrm{d}\mathbf{p}_k\right\}
$$



Basic geometry for tracking, inverse depth estimation and gradient estimation.
H. Kim, et. al, ECCV' 2016

Real-time 3D reconstruction and 6-DoF tracking with an event camera [ECCV 16]

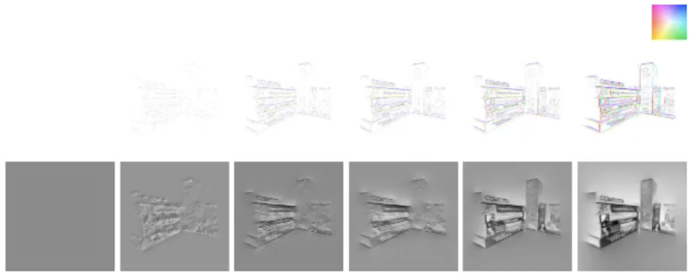
Filter2: Pixel-Wise EKF Based Gradient Estimation

Measurement Model

$$z_{\mathbf{g}} = \pm \frac{C}{\tau_c} ,$$
$$h_{\mathbf{g}} = (\mathbf{g}(\hat{\mathbf{p}}_k) \cdot \mathbf{m}) ,$$
$$\text{where } \mathbf{m} = \frac{\mathbf{p}_k^{(t)} - \mathbf{p}_k^{(t-\tau_c)}}{\tau_c}$$

Log Intensity Reconstruction

$$\min_{\mathrm{I}_l} \left\{ \int_{\Omega} \|\mathbf{g}(\mathbf{p}_k) - \nabla \mathrm{I}_l(\mathbf{p}_k)\|_{\epsilon_d}^{h} + \lambda \|\nabla \mathrm{I}_l(\mathbf{p}_k)\|_{\epsilon_r}^{h} \mathrm{d}\mathbf{p}_k \right\}$$



Temporal progression (left to right) of gradient estimation and log intensity reconstruction.

H. Kim, et. al, ECCV' 2016

Real-time 3D reconstruction and 6-DoF tracking with an event camera [ECCV 16]

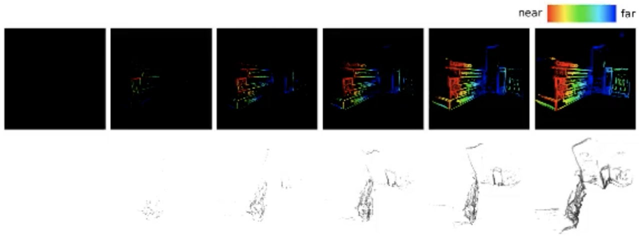
Filter3: Pixel-Wise EKF Based Inverse Depth Estimation

Measurement Model

$$z_{\rho} = \pm C ,$$
$$h_{\rho} = \mathrm{I}_l\left(\mathbf{p}_w^{(t)}\right) - \mathrm{I}_l\left(\mathbf{p}_w^{(t-\tau_c)}\right)$$

Inverse Depth Regularization

Penalises deviation from a spatially smooth inverse depth map by assigning each inverse depth value the average of its neighbours, weighted by their respective inverse variances.

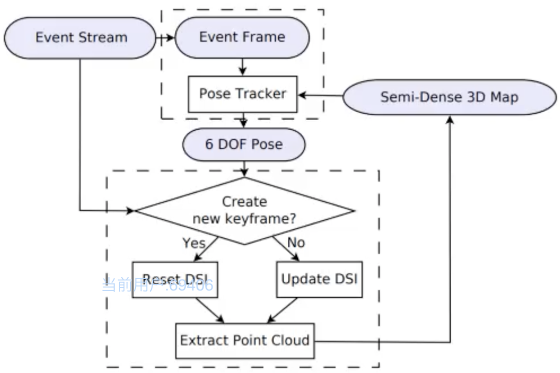


Temporal progression (left to right) of inverse depth estimation and regularization.

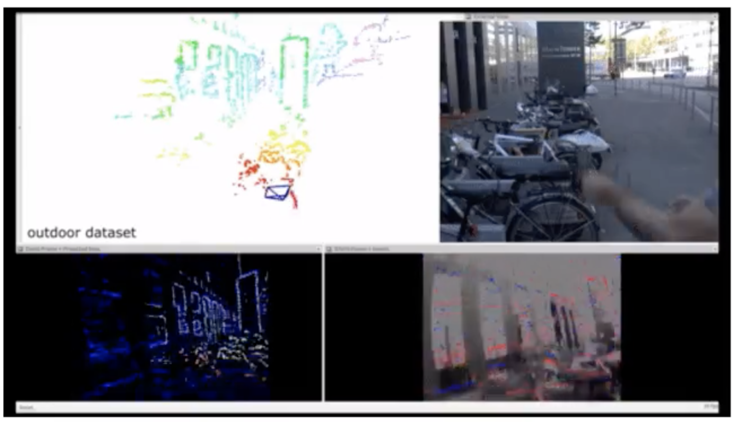
H. Kim, et. al, ECCV' 2016

EVO [RAL 17]

Pipeline Chart



H. Rebecq, et. al, RAL' 2017



Video courtesy:
https://www.youtube.com/watch?v=bYqD2qZJlxE&t=8s&ab_channel=UZHRoboticsandPerceptionGroup

EVO [RAL 17]

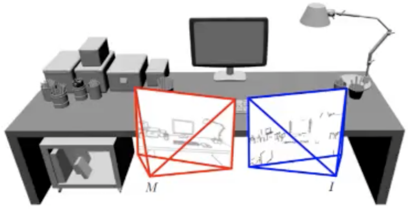
Tracking

Image-to-model alignment method

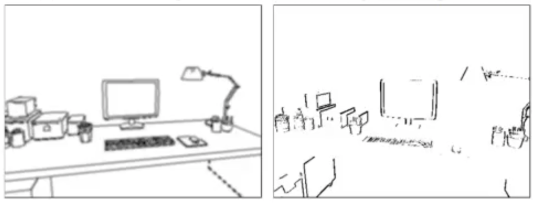
Registration is done using the inverse compositional Lucas Kanade (LK) method, by iteratively computing the incremental pose ΔT that minimizes

$$\sum_{\boldsymbol{u}} \left(M(\mathbf{W}(\boldsymbol{u};\Delta\mathrm{T})) - I(\mathbf{W}(\boldsymbol{u};\mathrm{T}))\right)^2$$

Event image I is obtained by aggregating a small number of events into an edge map.
Template M consists of the projected semi-dense 3D map of the scene according to a known pose of the event camera.



(a) 3D scene and poses involved in the registration process.



(b) Projected semi-dense map M (c) Event image I

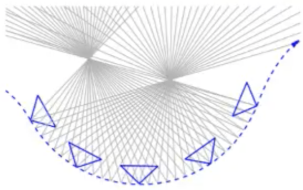
Pose Tracking computes the pose of the camera with respect to a reference pose by aligning the event image I with the projected semi-dense map M.

H. Rebecq, et. al., IEEE RAL, 2017

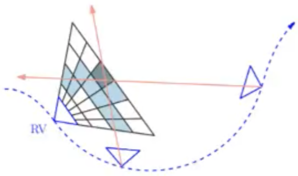
EVO [RAL 17]

Mapping

EMVS Pipeline

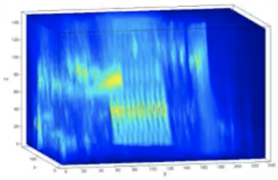


Back-projection of the events provides rays through space. The regions of high ray density mark the candidate locations of 3D edges.

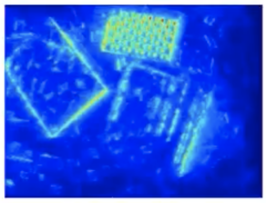


Space is discretized using a voxel grid centered at a virtual camera in a reference viewpoint. Each voxel value (blue) is the number of back-projected events (rays) traversing it.

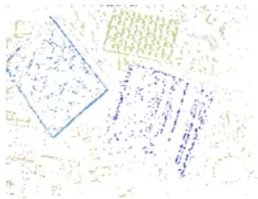
H. Rebecq, et. al., "EMVS: Event-based multi-view stereo—3D reconstruction with an event camera in real-time," IJCV. 2018.



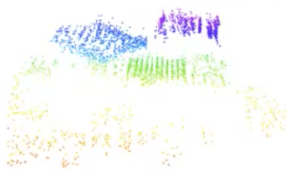
(a) Ray density DSI.



(b) Confidence map.



(c) 2D semi-dense depth map.



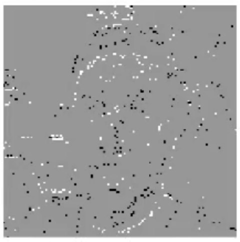
(d) 3D point cloud.

Mapping: the EMVS method builds the ray density DSI (a), from which a confidence map (b) and a semi-dense depth map (c) are extracted in a virtual camera. The semi-dense depth map gives a point cloud of scene edges (d). Images courtesy: H. Rebecq, et. al, IJCV' 2018.

How to unlock event cameras' potential

What is NOT event based Machine Vision

embedded VISION SUMMIT 2018



(a) Event Data

(b) Image Reconstruction

Do not generate Images from Events

- Not event based,
- Useless approach, GPU use, to generate 100-200Hz bad quality images
- Fake SLAM, using binary images....

C. Reinbacher, G. Graber, T. Pock, Real-Time Intensity-Image Reconstruction for Event Cameras Using Manifold Regularisation, British Machine Vision Conference, 2016

Slide courtesy: a Presentation from Ryad B. Benosman

Data Association

Classical Method
(using standard cameras)

- ☐ Feature Correspondence (PTAM, ORB-SLAM)
- ☐ Photometric Consistency (LSD-SLAM, DSO)
- ☐ Geometric Distance (KinectFusion, CannyVO)

Event-based Method

- ☐ 6-DoFf Pose Tracking
- ☐ Log Intensity Gradient
- ☐ Inverse Depth Estimation

→ Brightness constancy (Log domain)
Linear gradient assumption

H. Kim, et. al, ECCV' 2016

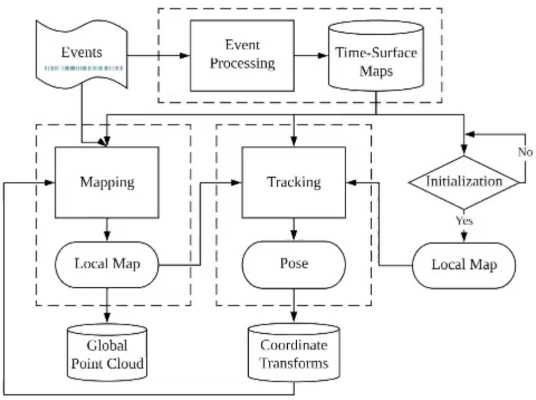
- ☐ 6-DoFf Pose Tracking → 3D-2D point-to-edge distance
- ☐ Inverse Depth Estimation → Ray density

H. Rebecq, et. al., IEEE RAL, 2017

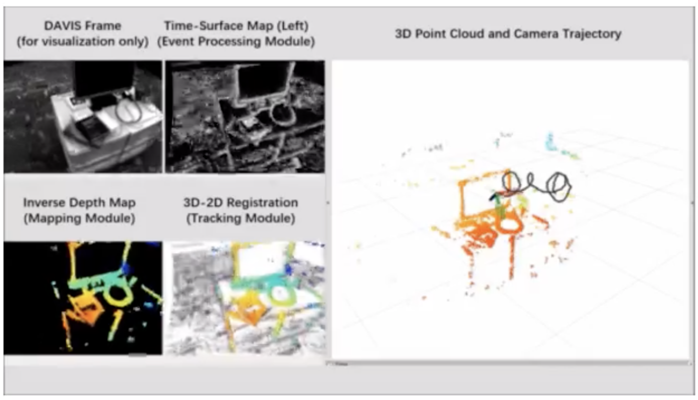
Explore Novel Ideas

- ☐ True event-based methods that can unlock event cameras' potential.
- ☐ Can we find new X-metric information based on which the event-based data association is established?
- ☐ Is the monocular configuration the best choice? (How about stereo?)
- ☐ ...

Event-based Stereo Visual Odometry (ESVO)



System flowchart.



Home Page: https://sites.google.com/view/esvo-project-page/home

Time-Surface Map

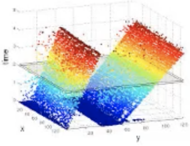
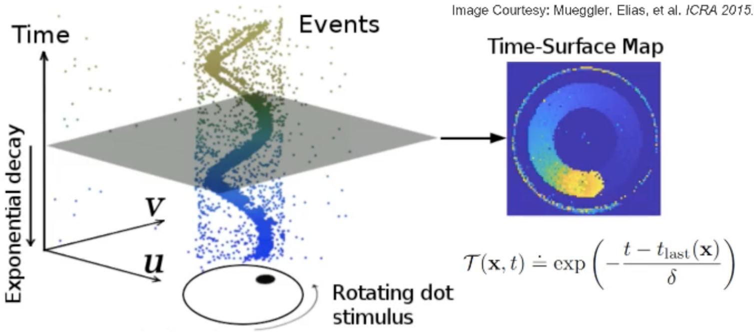


Image Courtesy: Mueggler. Elias, et al. *ICRA 2015.*



Event Representation

[1] Lagorce, X., *et al*, R.: HOTS: a hierarchy of event-based time-surfaces for pattern recognition. IEEE Trans. Pattern Anal. Mach. Intell. 2016

Time-Surface Map

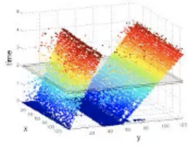
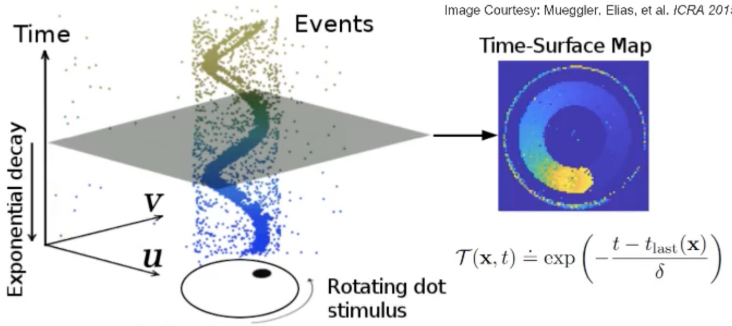
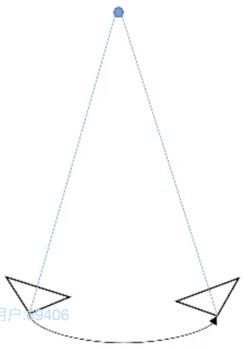


Image Courtesy: Mueggler. Elias, et al. *ICRA 2015.*

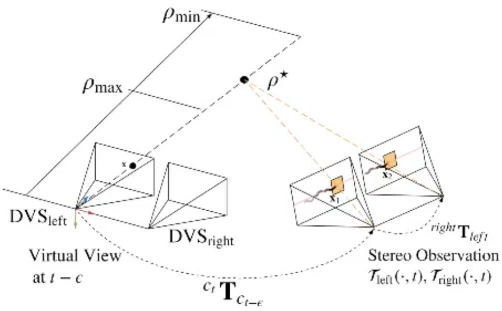


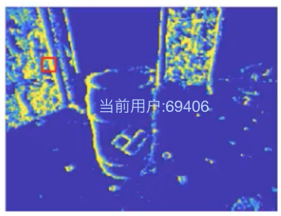
Event Representation

[1] Lagorce, X., *et al*, R.: HOTS: a hierarchy of event-based time-surfaces for pattern recognition. IEEE Trans. Pattern Anal. Mach. Intell. 2016

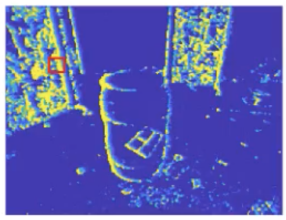


Spatio-Temporal Consistency (Data Association)



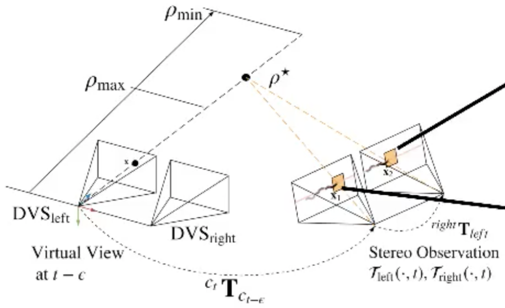


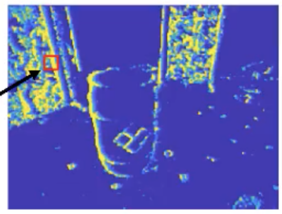
Time-Surface Map (Right)



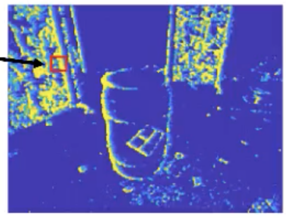
Time-Surface Map (Left)

Spatio-Temporal Consistency (Data Association)





Time-Surface Map (Right)

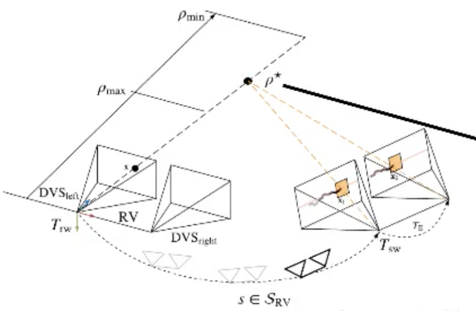


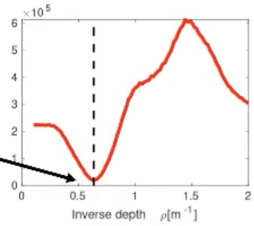
Time-Surface Map (Left)

$$\rho^\star = \arg\min_{\rho} C(\mathbf{x}, \rho, \mathcal{T}_{\text{left}}(\cdot,t), \mathcal{T}_{\text{right}}(\cdot,t), \mathbf{T}_{t-\delta t:t})$$

$$C(\cdots) \doteq \sum_{\mathbf{x}_{1,i}\in W_1, \mathbf{x}_{2,i}\in W_2} \left\|\tau^t_{\text{left}}(\mathbf{x}_{1,i}) - \tau^t_{\text{right}}(\mathbf{x}_{2,i})\right\|_2^2.$$

Inverse Depth Estimation





Objective Function

Algorithm 1 Inverse Depth Estimation

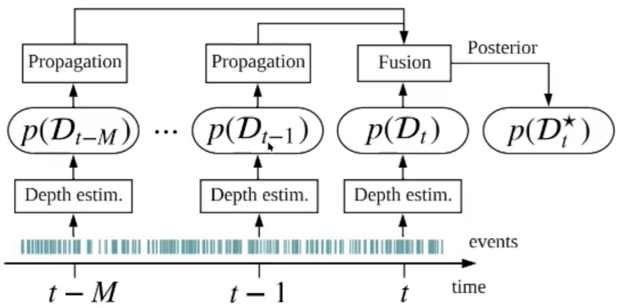
1: **Input**: event $e_{t-\epsilon}$, stereo event observation $\mathcal{T}^t_{\text{left}}, \mathcal{T}^t_{\text{right}}$ and the relative transformation ${}^{c_t}\mathbf{T}_{c_{t-\epsilon}}$.
2: Initialize ρ by ZNCC-block matching on $\mathcal{T}^t_{\text{left}}, \mathcal{T}^t_{\text{right}}$.
3: **while** not converged **do**
4:     Compute residuals $\mathbf{r}(\rho)$ in (4).
5:     Compute Jacobian $\mathbf{J}(\rho)$ (analytical derivatives).
6:     Update: $\rho \leftarrow \rho + \Delta\rho$, using (7).
7: **end while**
8: **return** Inverse depth ρ.

$$\rho^\star = \arg\min_{\rho} C(\mathbf{x}, \rho, \mathcal{T}_{\text{left}}(\cdot,t), \mathcal{T}_{\text{right}}(\cdot,t), \mathbf{T}_{t-\delta t:t})$$

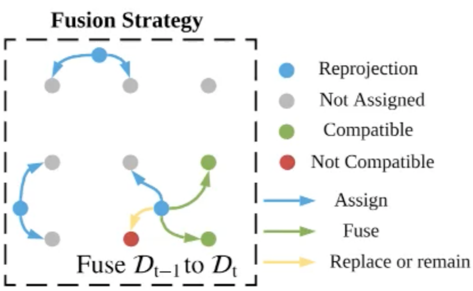
$$C(\cdots) \doteq \sum_{\mathbf{x}_{1,i}\in W_1, \mathbf{x}_{2,i}\in W_2} \left\|\tau^t_{\text{left}}(\mathbf{x}_{1,i}) - \tau^t_{\text{right}}(\mathbf{x}_{2,i})\right\|_2^2.$$

Semi-Dense Reconstruction

Inverse Depth Propagation and Fusion



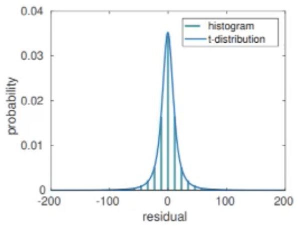
(a) Flowchart of mapping module.



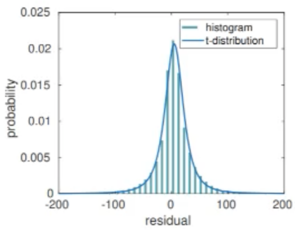
(b) Depth fusion rules at locations on a 3 × 3 pixel grid.

Semi-Dense Reconstruction

Probabilistic Characteristics



(a) *simulation_3planes* [59].



(b) *upenn_flying1* [56].

Student's t Distribution

$$\mathbf{x} \sim St(\boldsymbol{\mu}, \boldsymbol{S}, \nu)$$

$$\mathbf{z} = \mathbf{Ax} + \mathbf{b} \sim St(\mathbf{A}\boldsymbol{\mu} + \mathbf{b},\, \mathbf{A}\boldsymbol{S}\mathbf{A}^\top,\, \nu)$$

Propagation

$$\Delta\rho \sim St\left(-\frac{\sum J_i}{\|\mathbf{J}\|^2}\mu_r,\ \frac{s_r^2}{\|\mathbf{J}\|^2},\ \nu_r\right)$$

$$\rho \sim St\left(\rho^\star,\ \frac{s_r^2}{\|\mathbf{J}\|^2},\ \nu_r\right)$$

$$\sigma_{\rho^\star}^2 = \frac{\nu_r}{\nu_r - 2}\,\frac{s_r^2}{\|\mathbf{J}\|^2}$$

Update

$$\nu' = \min(\nu_a, \nu_b),$$

$$\mu = \frac{\sigma_a^2\mu_b + \sigma_b^2\mu_a}{\sigma_a^2 + \sigma_b^2},$$

$$\sigma^2 = \frac{\nu' + \frac{(\mu_a - \mu_b)^2}{\sigma_a^2 + \sigma_b^2}}{\nu' + 1} \cdot \frac{\sigma_a^2\sigma_b^2}{\sigma_a^2 + \sigma_b^2},$$

$$\nu = \nu' + 1.$$

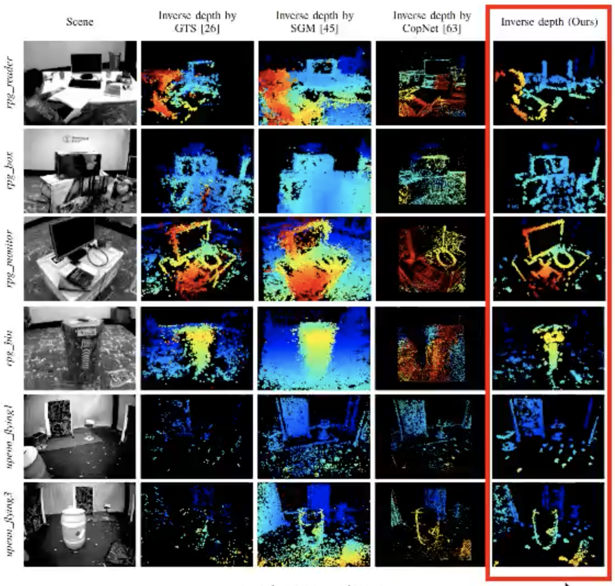
Mapping Results

TABLE IV: Quantitative evaluation of mapping on sequences with ground truth depth.

	Sequence [56]	*upenn_flying1*	*upenn_flying3*
	Depth range [m]	5.48 m	6.03 m
GTS [26]	Mean error	0.31 m	0.44 m
	Median error	0.18 m	0.21 m
	Relative error	5.64 %	7.26 %
SGM [45]	Mean error	0.31 m	0.20 m
	Median error	0.15 m	0.10 m
	Relative error	5.58 %	3.28 %
CopNet [63]	Mean error	0.59 m	0.53 m
	Median error	0.49 m	0.44 m
	Relative error	10.93 %	8.87 %
Our Method	Mean error	**0.16** m	**0.19** m
	Median error	**0.12** m	**0.09** m
	Relative error	**3.05** %	**3.13** %

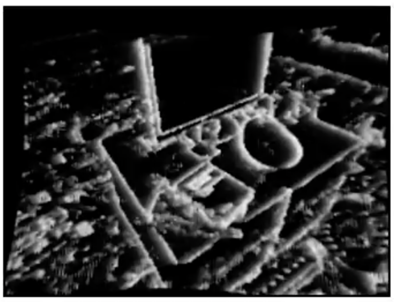
Good Reference Map for Tracking

- Sharp, Semi-Dense Edge Map
- Noise-Free



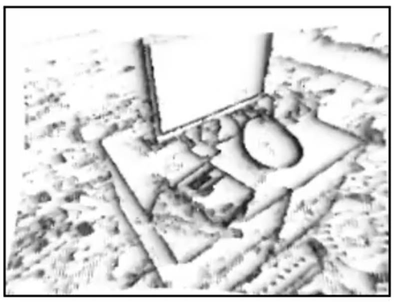
Qualitative Evaluation.

Exploiting Time Surfaces as Distance Fields



Time Surface $\tau(\mathbf{x})$

$$\bar{\tau}(\mathbf{x}) = 1 - \tau(\mathbf{x})$$



Time Surface Negative $\bar{\tau}(\mathbf{x})$

Anisotropic Distance Field

Inspiration: Y. Zhou, *et. al.* "Canny-vo: Visual odometry with rgb-d cameras based on geometric 3-d–2-d edge alignment." *IEEE Transactions on Robotics* 35.1 (2018): 184-199.

3D-2D Registration

Objective Function

$$\boldsymbol{\theta}^{\star} = \arg\min_{\boldsymbol{\theta}} \sum_{\mathbf{x}\in\mathcal{D}^{\mathcal{F}_{\text{ref}}}} \|\bar{\boldsymbol{\tau}}^{\mathcal{F}_k}_{\text{left}}(W(\mathbf{x},\rho;\boldsymbol{\theta}))\|^2$$

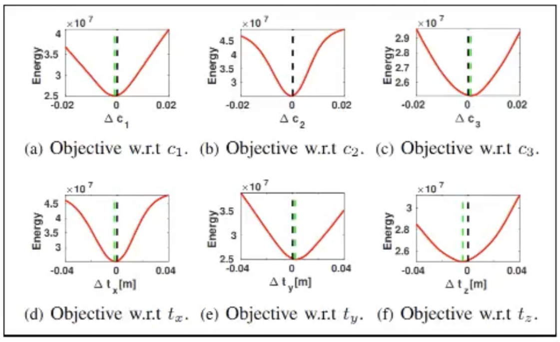
$$W(\mathbf{x},\rho;\boldsymbol{\theta}) \doteq \pi_{\text{left}}(T(\pi^{-1}_{\text{ref}}(\mathbf{x},\rho), G(\boldsymbol{\theta})))$$

$$\boldsymbol{\theta} \doteq [c_1, c_2, c_3, t_x, t_y, t_z]^T, \quad G(\boldsymbol{\theta}) : \mathbb{R}^6 \to \mathrm{SE}(3)$$

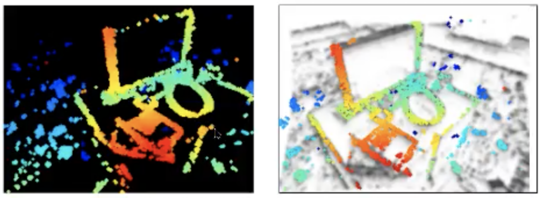
Forward Compositional LK Method

$$F(\Delta\boldsymbol{\theta}) \doteq \sum_{\mathbf{x}\in\mathcal{D}^{\mathcal{F}_{\text{ref}}}} \| \underbrace{\bar{\boldsymbol{\tau}}^{\mathcal{F}_k}_{\text{left}}(W(W(\mathbf{x},\rho;\Delta\boldsymbol{\theta});\boldsymbol{\theta}))}_{r_{\mathbf{x}}} \|^2$$

$$W(\mathbf{x},\rho;\boldsymbol{\theta}) \leftarrow W(\mathbf{x},\rho;\boldsymbol{\theta}) \circ W(\mathbf{x},\rho;\Delta\boldsymbol{\theta})$$



Slices of the objective function.



(a) Inverse depth map in the reference frame. (b) Registration on the negative time-surface map.

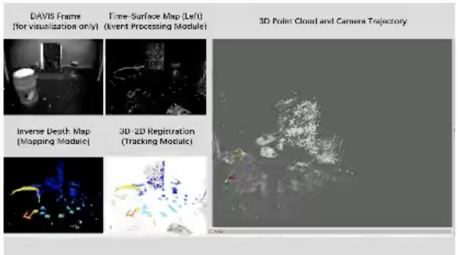
Proposed tracking method.

Results

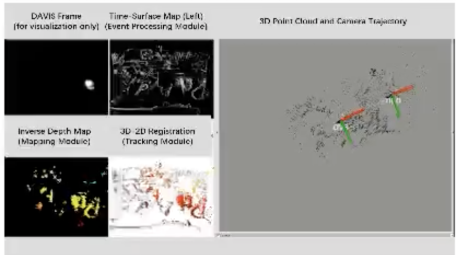
DAVIS 346 Stereo Rig

TABLE VII: Computational performance

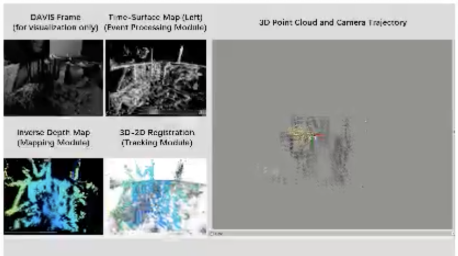
Node (#Threads)	Function	Time (ms)
Time surfaces (1)	Exponential decay	5 – 10
Initialize depth (1)	SGM & masking	12 – 17
Mapping (4)	Event matching	6 (~ 1000 events)
	Depth optimization	15 (~ 500 events)
	Depth fusion	20 (~ 60000 fusions)
Tracking (2)	Non linear solver	10 (300 points × 5 iterations)



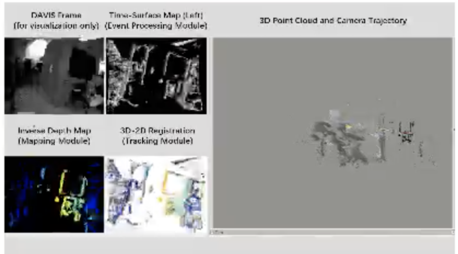
Upenn_flyingroom (Drone)



HKUST HDR (Handheld)



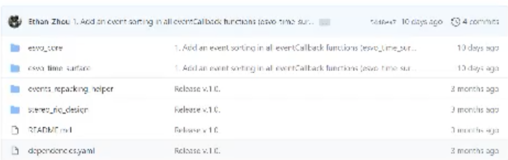
UZH rpg (Handheld)

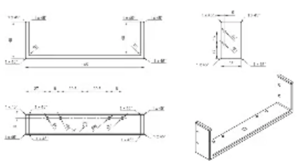


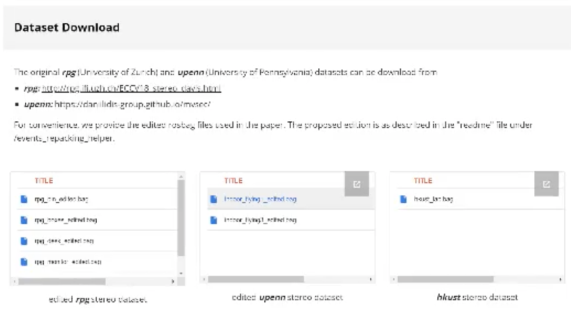
HKUST Lab (Handheld)

Open-Source Project

- Home Page: https://sites.google.com/view/esvo-project-page/home
- PDF: https://arxiv.org/pdf/2007.15548.pdf
- Code: https://github.com/HKUST-Aerial-Robotics/ESVO







Recap Event Cameras' Advantages

- High dynamic range (HDR)
- High speed
- Low latency
- Low power consumption

Limitations

- Low spatial resolution
- Bad signal-noise ratio (sometimes)
- Expensive

Challenges

SLAM Perspective

- Needs more dedicated benchmark dataset
- Sensor suite
- ...

General Research Perspective

- Proper hardware
- Other interesting problems (*e.g., pattern recognition, behavior inference, etc*)
- Learning with event camera (*e.g.* ANN, SNN)
- ...

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Thanks You!

Time for questions!

Yi Zhou
eeyzhou@ust.hk

2020.12.02