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

Event-based Visual Odometry: A Short Tutorial

Dr. Yi Zhou
eeyzhou@ust.hk

HKUST-DJI Joint Lab,
Dept. ECE at HKUST

About Event-based Cameras

Working principle:

◆Asynchronous and independent pixels

Properties:

◆High speed, low latency (~ 1 μs)

◆High dynamic range (140 dB instead of 60 dB)

◆Ultra-low power (mean: 1mW)

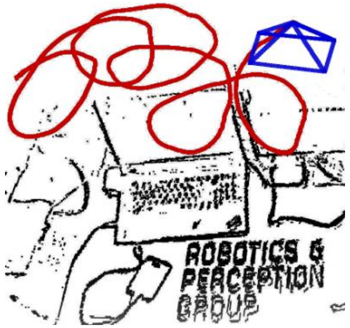
Challenges

Event streams cannot be fed directly to existing methods designed for standard cameras!

Question to Answer

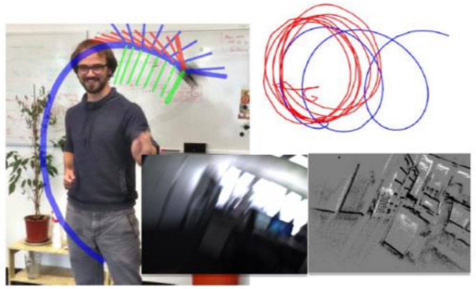
"How to leverage the advantages of event-based cameras to solve a given task by optimally processing the event stream?"

Event-based VO

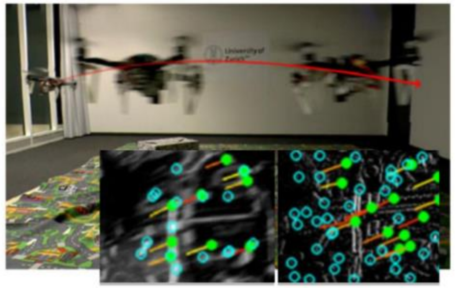


EVO [RAL 17]

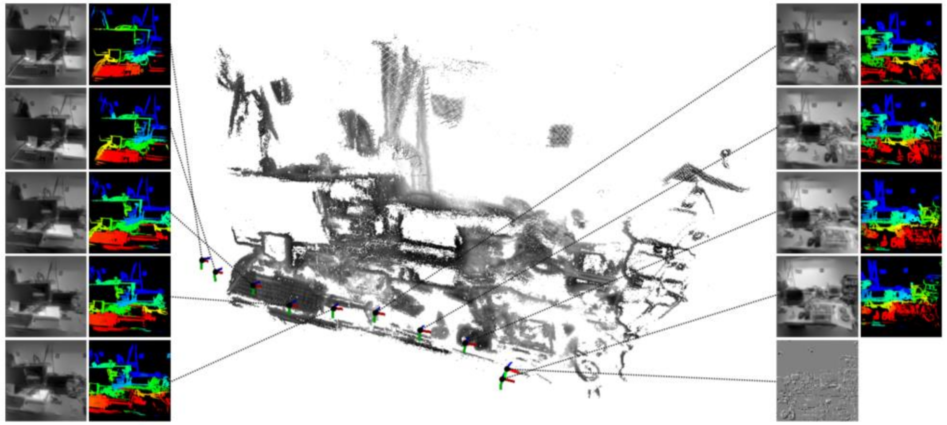
G. Gallego *et. al* [T-PAMI 18]



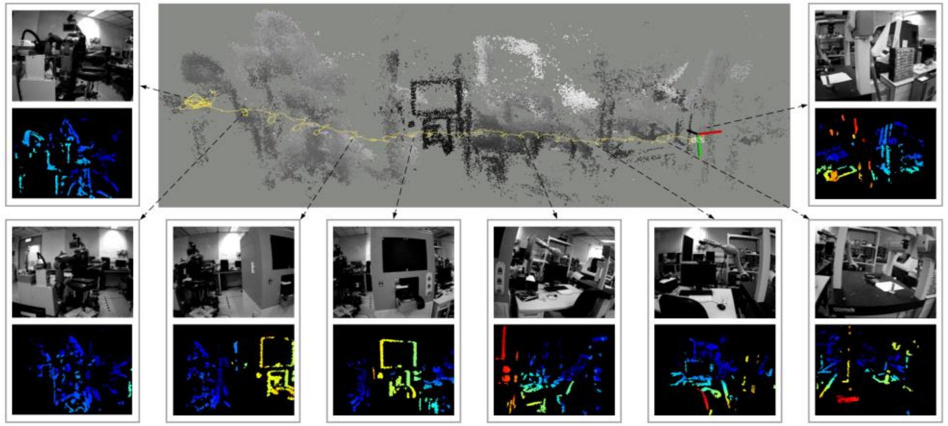
H. Rebecq [*BMVC 17*]



Ultimate SLAM [*RAL 18*]



H. Kim [*ECCV 16*]

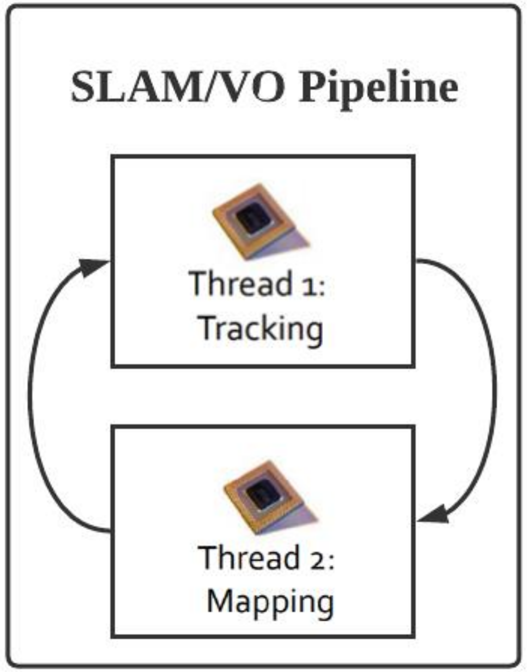


ESVO [*T-RO 21*]

Outline

1. A literature review
2. An introduction to ESVO system
3. Some take-home messages

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*]

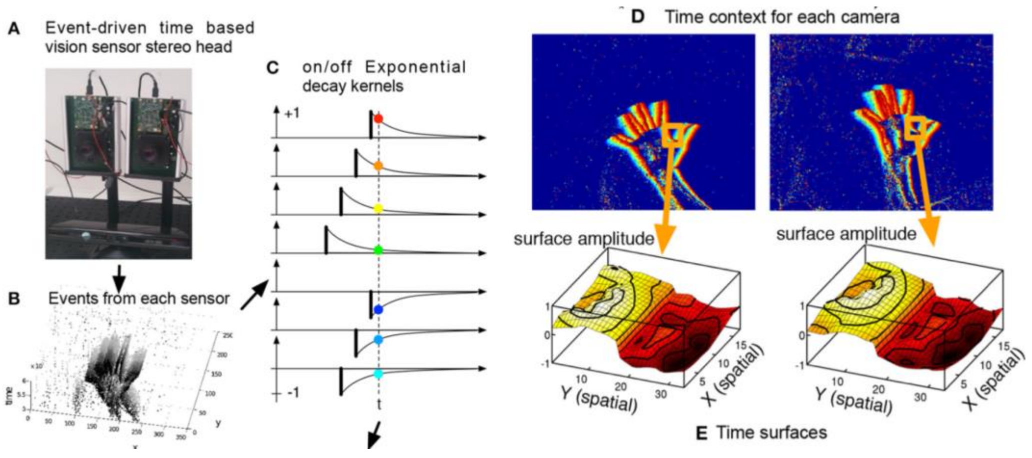
☐ Event-based VO Systems

[*ECCV 16, RAL 17*]

Event-based Mapping (3D Reconstruction)

☐ Instantaneous Stereo

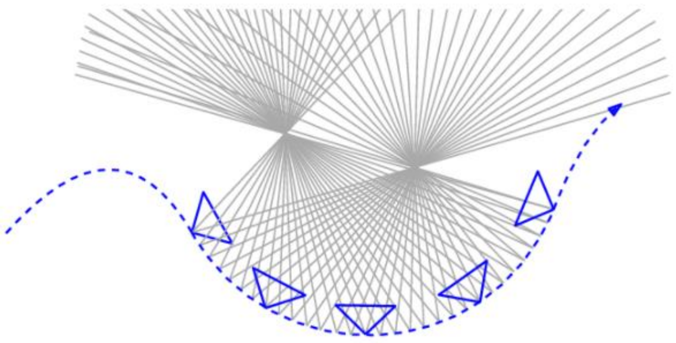
Two-Step paradigm
① Finding eipolar matching
② Triangulation



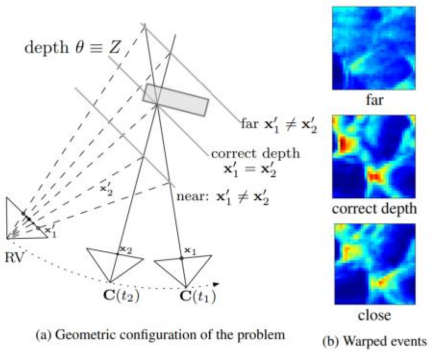
SH. Ieng, *et. al.*, Neuromorphic Event-Based Generalized Time-Based Stereovision, Front. Neurosci. 2018

☐ Temporal Stereo (monocular event camera!)

① Require prior knowledge of the camera's motion
② Use occurred over a temporal window



[*IJCV 18*]



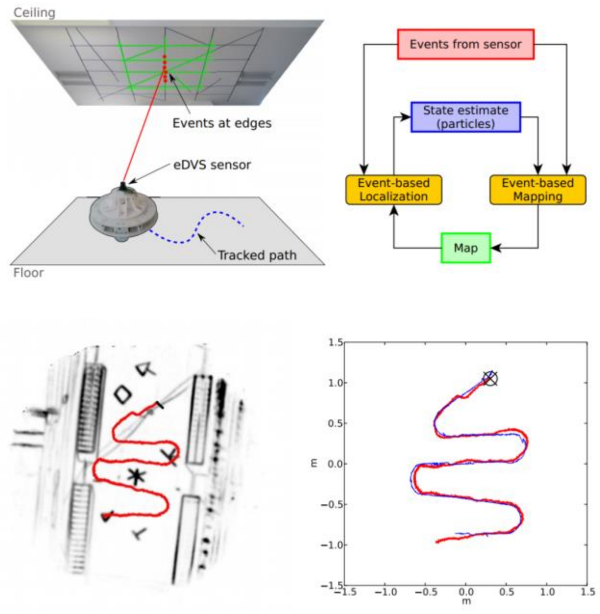
[*CVPR 18*]

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 Tracking

Motion and Scene Complexity：Simple -> Complex

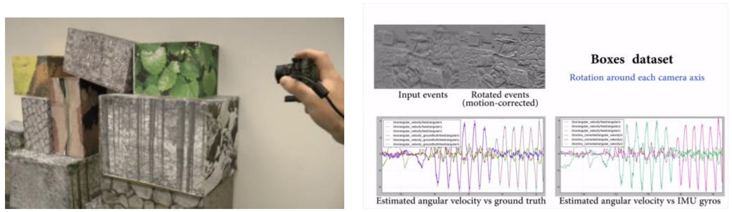
Planar Motion



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

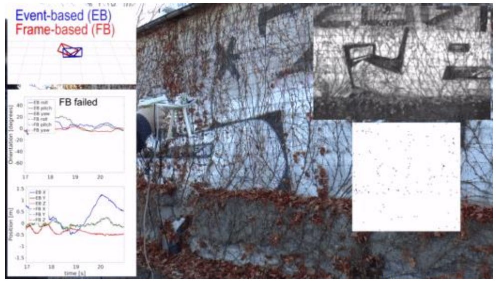
3D Rotation

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

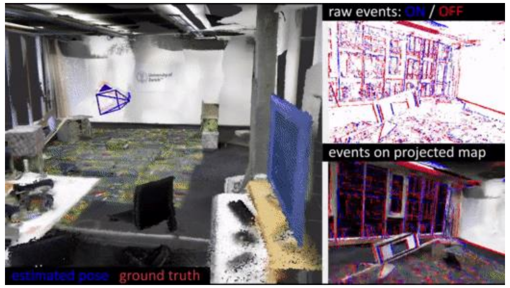


G. Gallego, et. al., “Accurate angular velocity estimation with an event camera,” RAL 2017

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 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.

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 RA-L, 2017.

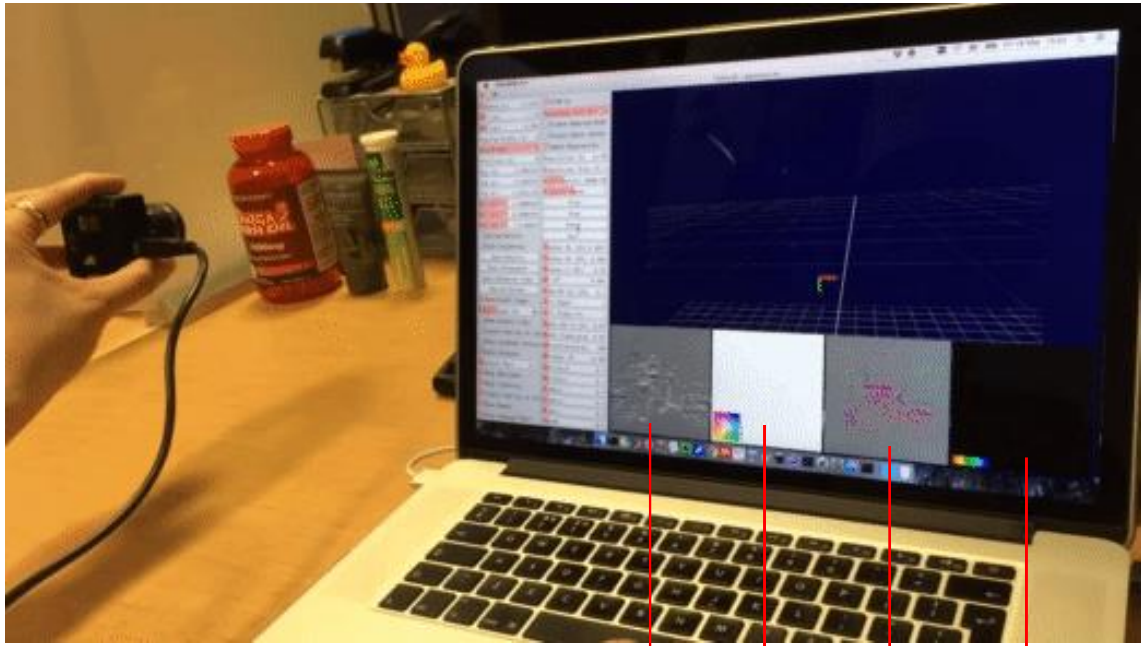
Y. Zhou, G. Gallego, and S. Shen. "Event-based stereo visual odometry (ESVO)." IEEE Transactions on Robotics, 2021.
(Project page: https://sites.google.com/view/esvo-project-page/home)

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

Method Outline

Three interleaved probabilistic filters (EKFs)

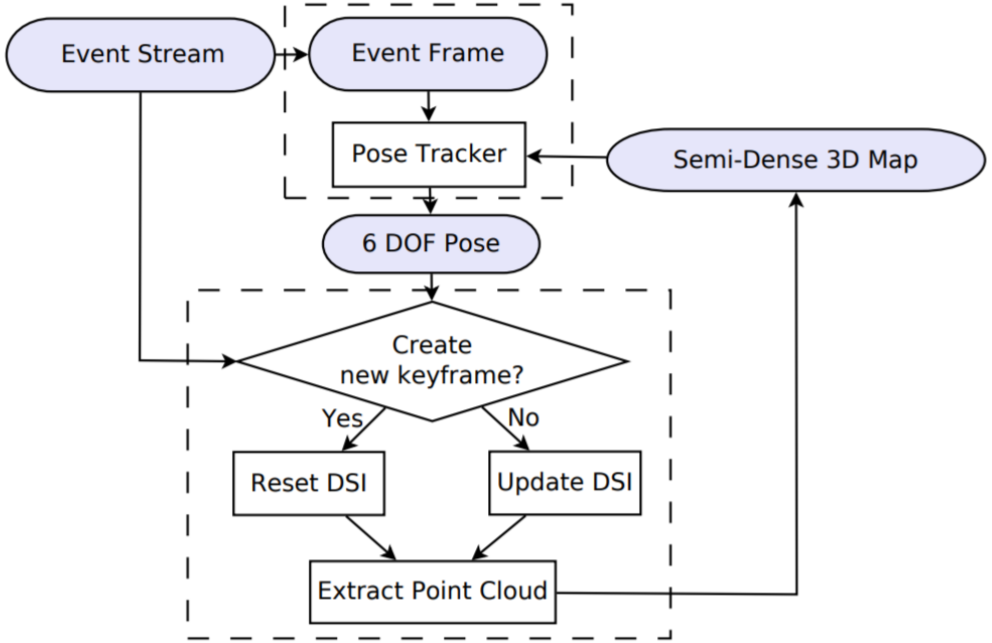
- Filter 1: Tracks global 6-DoF camera motion
- Filter 2: Estimates the log intensity gradients in a keyframe image
- Filter 3: Estimates the inverse depths of a keyframe



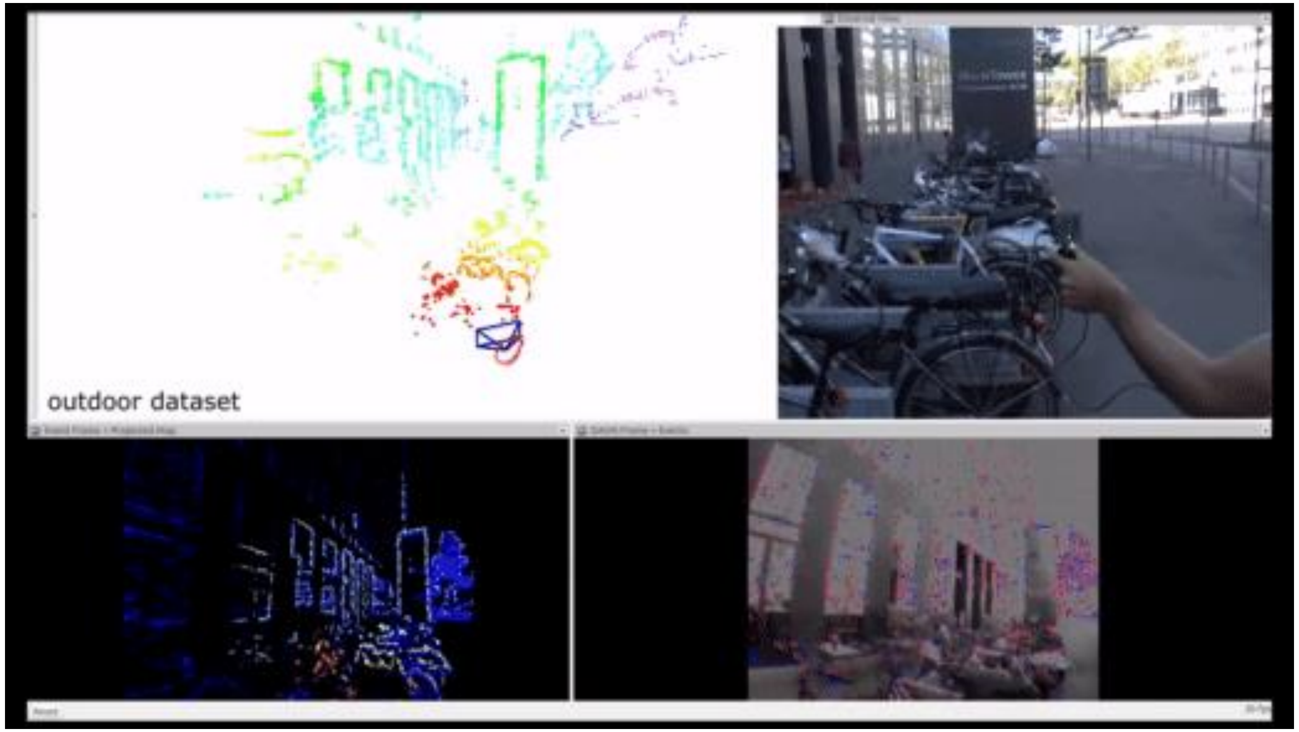
Video courtesy:
https://www.youtube.com/watch?v=yHLyhdMSw7w&ab_channel=HanmeKim

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

Core Problem of Event-based VO

VO Problem $p(\mathbf{x}_t, \mathbf{m}_t | \mathbf{z}_t)$ Recursive State Estimation

Prediction
$$p(\mathbf{x}_t, \mathbf{m}_t | \mathbf{z}_{t-1}) = \int \int p(\mathbf{x}_t, \mathbf{m}_t | \mathbf{x}_{t-1}, \mathbf{m}_{t-1}) \times p(\mathbf{x}_{t-1}, \mathbf{m}_{t-1} | \mathbf{z}_{t-1}) d\mathbf{x}_{t-1} d\mathbf{m}_{t-1}$$

Correction
$$p(\mathbf{x}_t, \mathbf{m}_t | \mathbf{z}_t) \propto p(\mathbf{z}_t | \mathbf{x}_t, \mathbf{m}_t) \times p(\mathbf{x}_t, \mathbf{m}_t | \mathbf{z}_{t-1})$$

Implementation Perspective

Tracking Subproblem $p(\mathbf{x}_t | \mathbf{z}_t, \mathbf{m}_t^*)$

$\mathbf{x}_t^*$ $\mathbf{m}_t^*$

Mapping Subproblem $p(\mathbf{m}_t | \mathbf{z}_t, \mathbf{x}_t^*)$

Core Problem of State Estimation from a Methodology Perspective

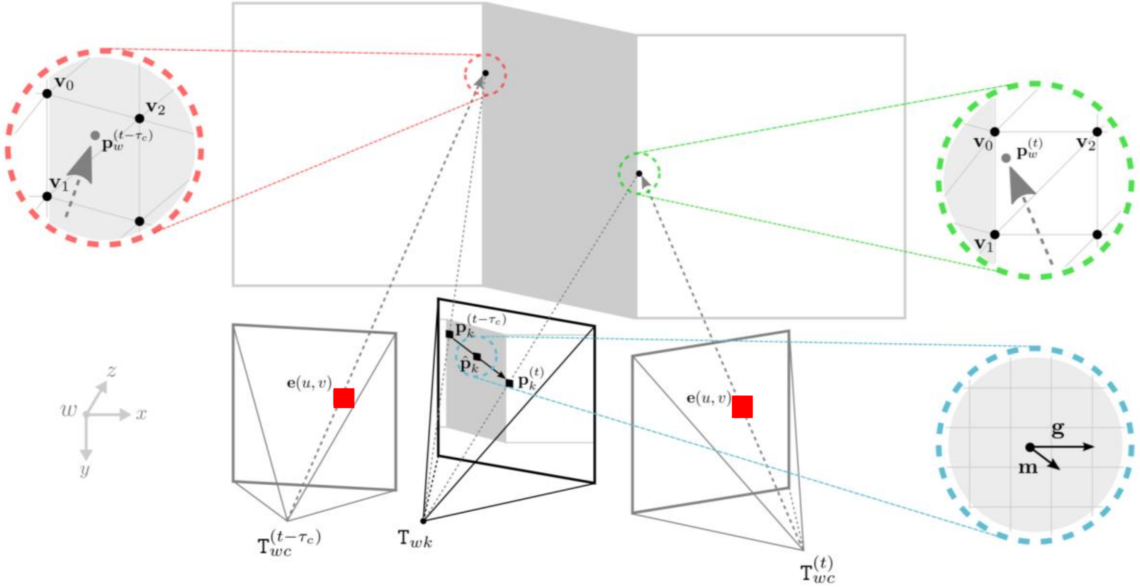
- ☐ Data Association
- ☐ Measurement Model

Core Problem of Event-based VO

Filter1: Tracks 6-DoF camera motion

Measurement Model

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



Filter2: Pixel-Wise EKF Based Gradient Estimation

Measurement Model

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

Filter3: Pixel-Wise EKF Based Inverse Depth Estimation

Measurement Model

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

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.

Core Problem of Event-based VO

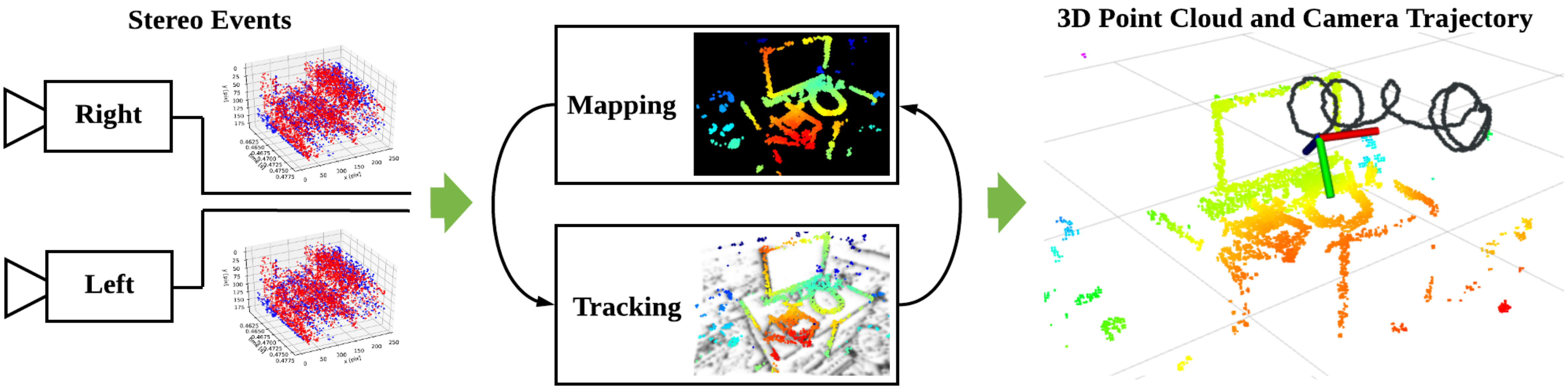
How to Make A Difference?

- ☐ Can we find a novel X-metric information based on which the event-based data association is established?
- ☐ Is the monocular configuration the best choice? (How about stereo?)
- ☐ …

ESVO: Event-based Stereo Visual Odometry

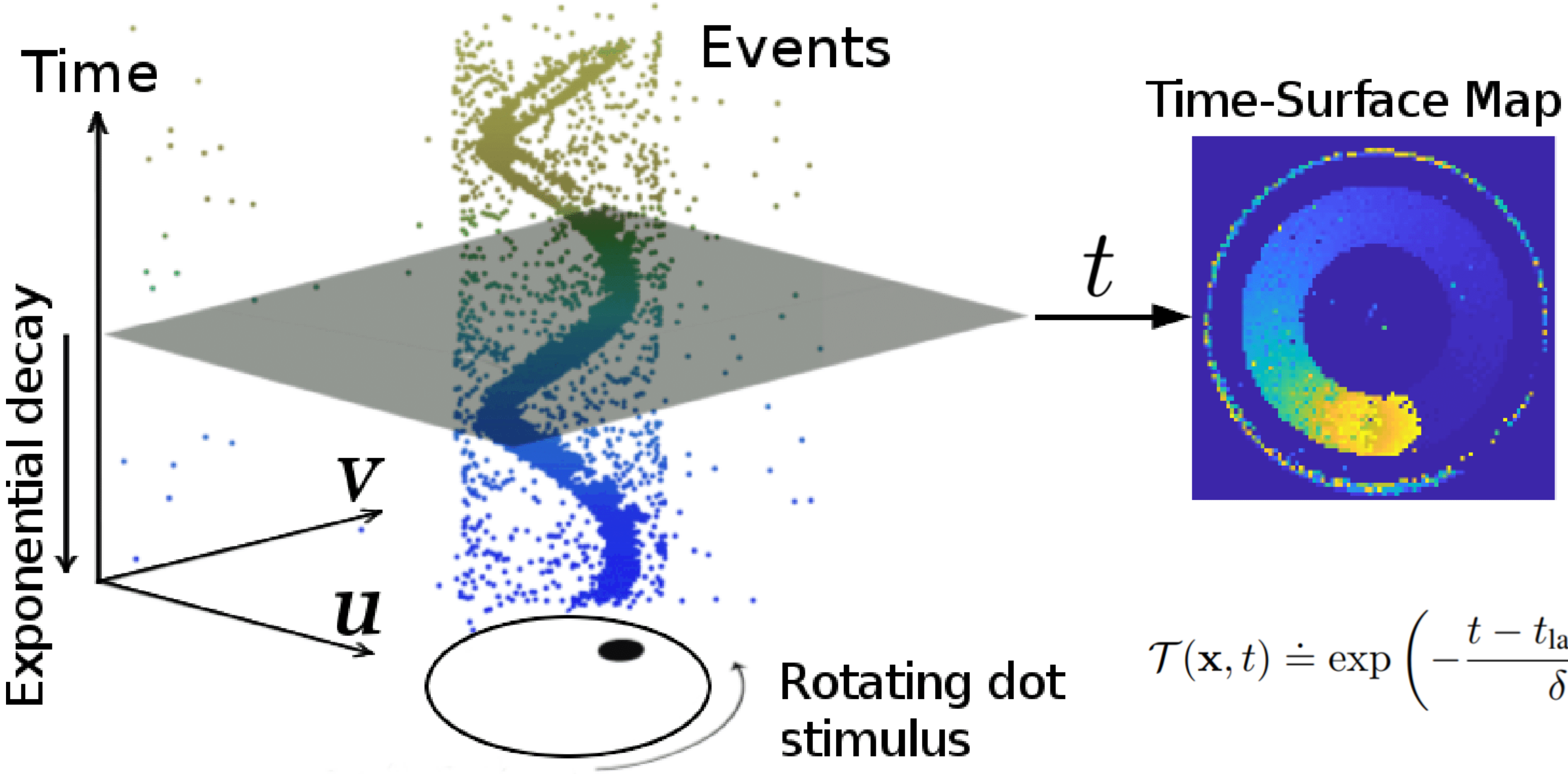

Stereo Event-based
Camera Rig

Our System


Stereo Events
Right
Left
Mapping
Tracking
3D Point Cloud and Camera Trajectory

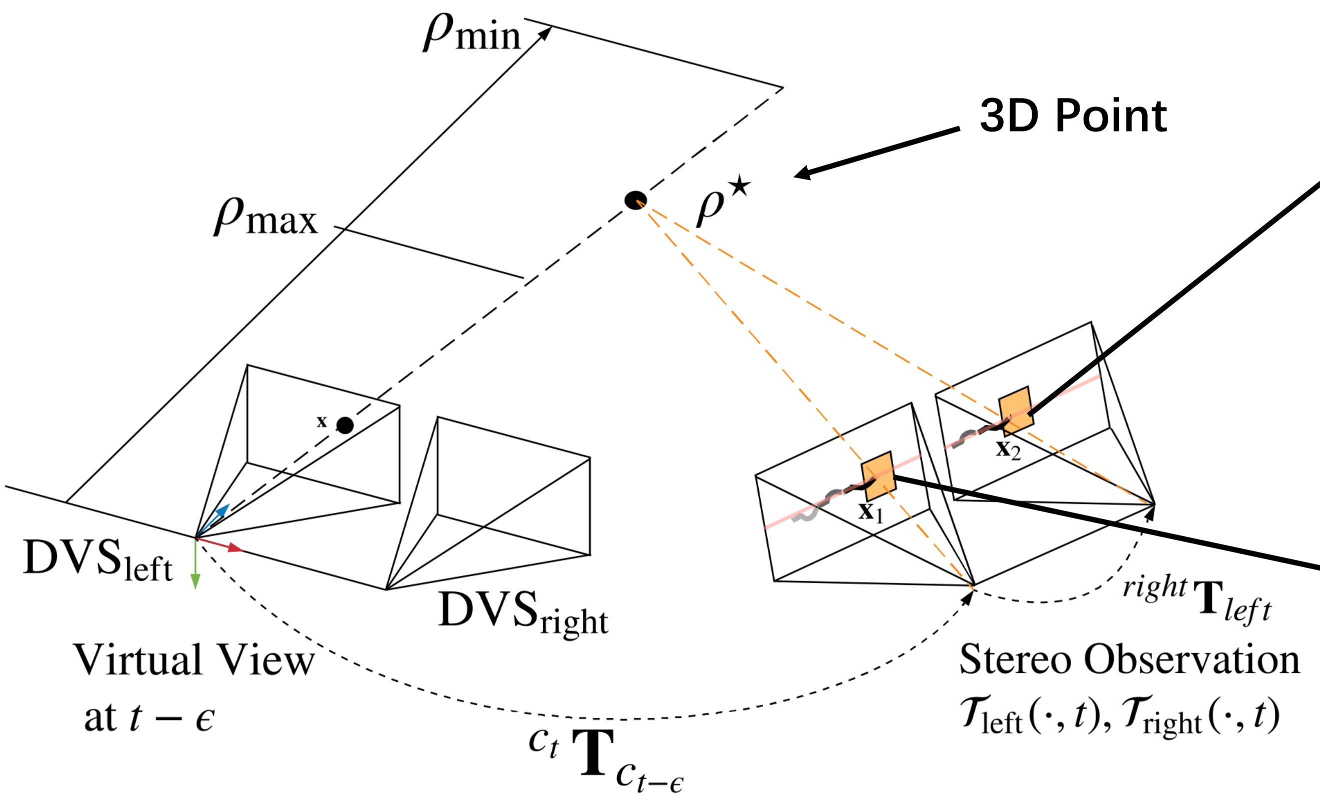
Mapping

Time-Surface Map



$$\mathcal{T}(\mathbf{x}, t) \doteq \exp\left(-\frac{t - t_{\text{last}}(\mathbf{x})}{\delta}\right)$$

Geometry



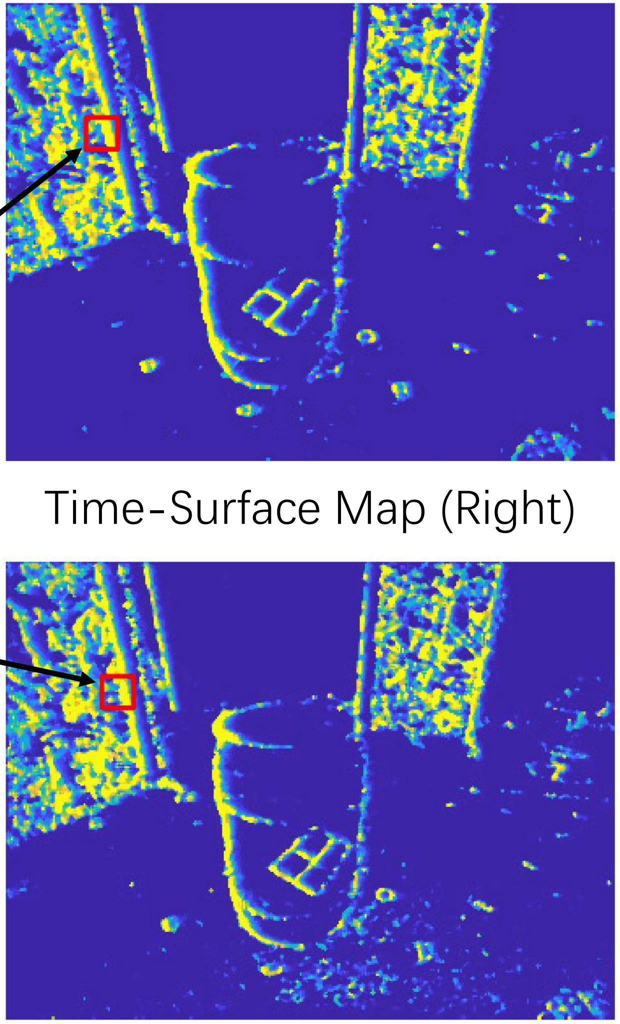
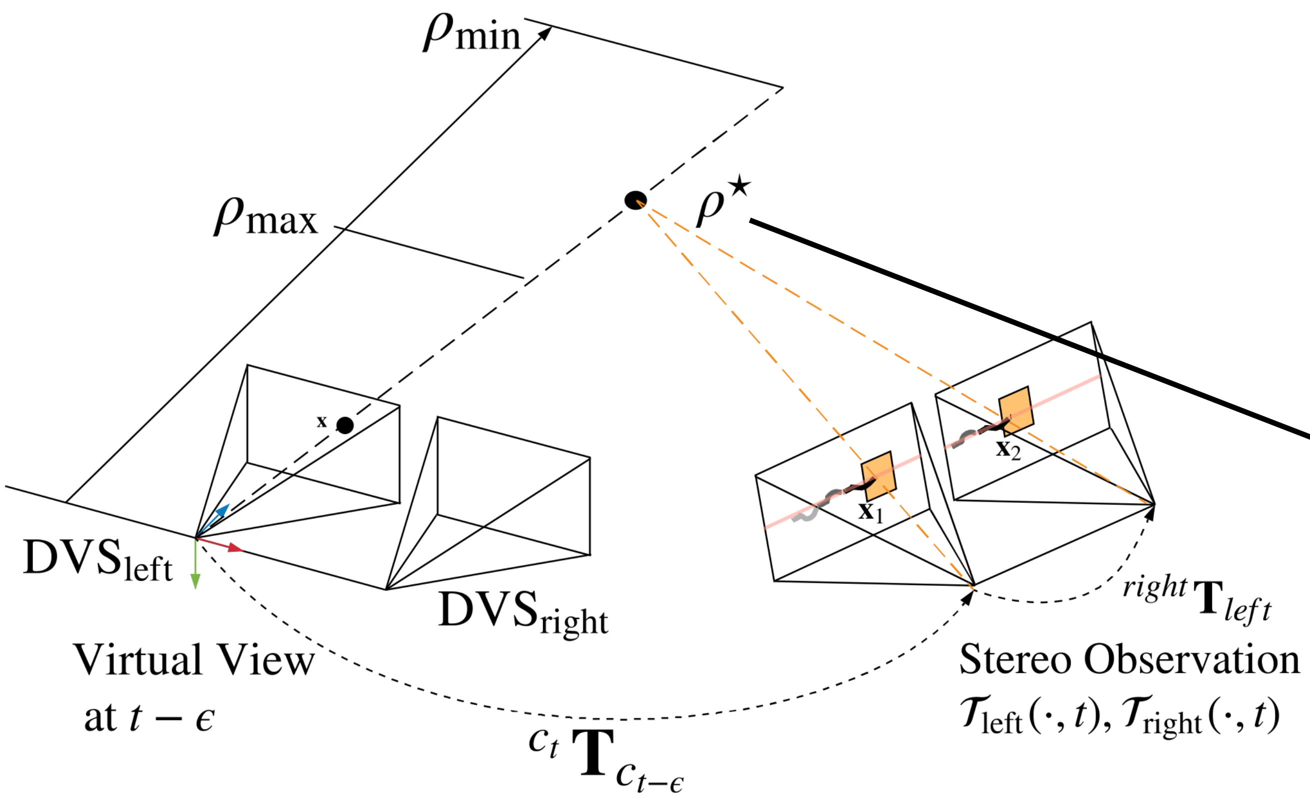
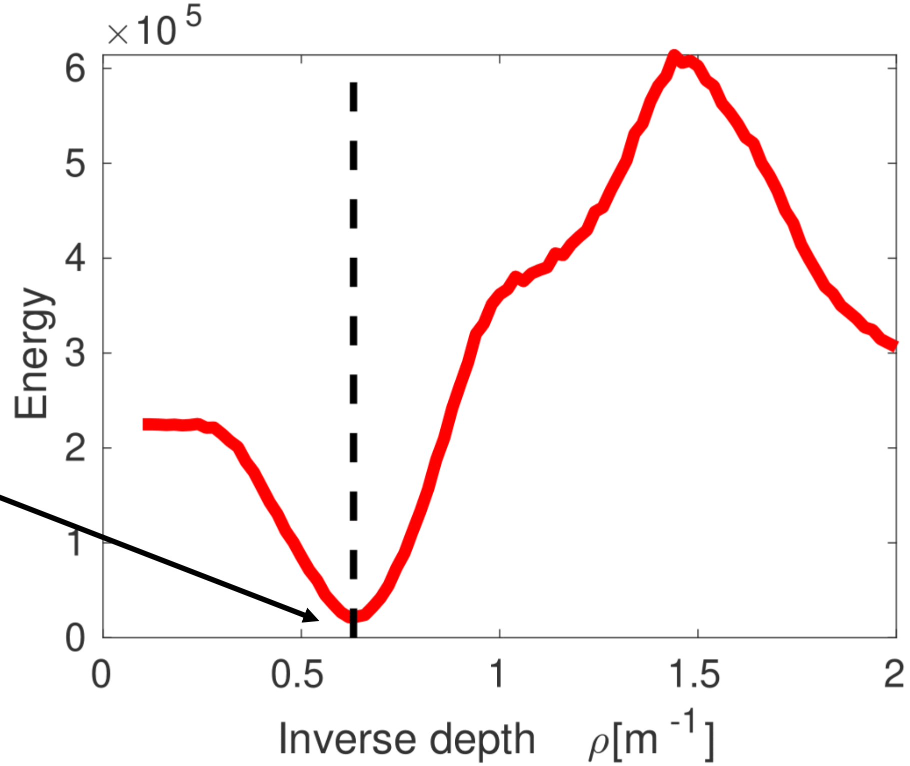


Illustration of the geometry of the proposed mapping method.

Time-Surface Map (Left)

Problem Formulation



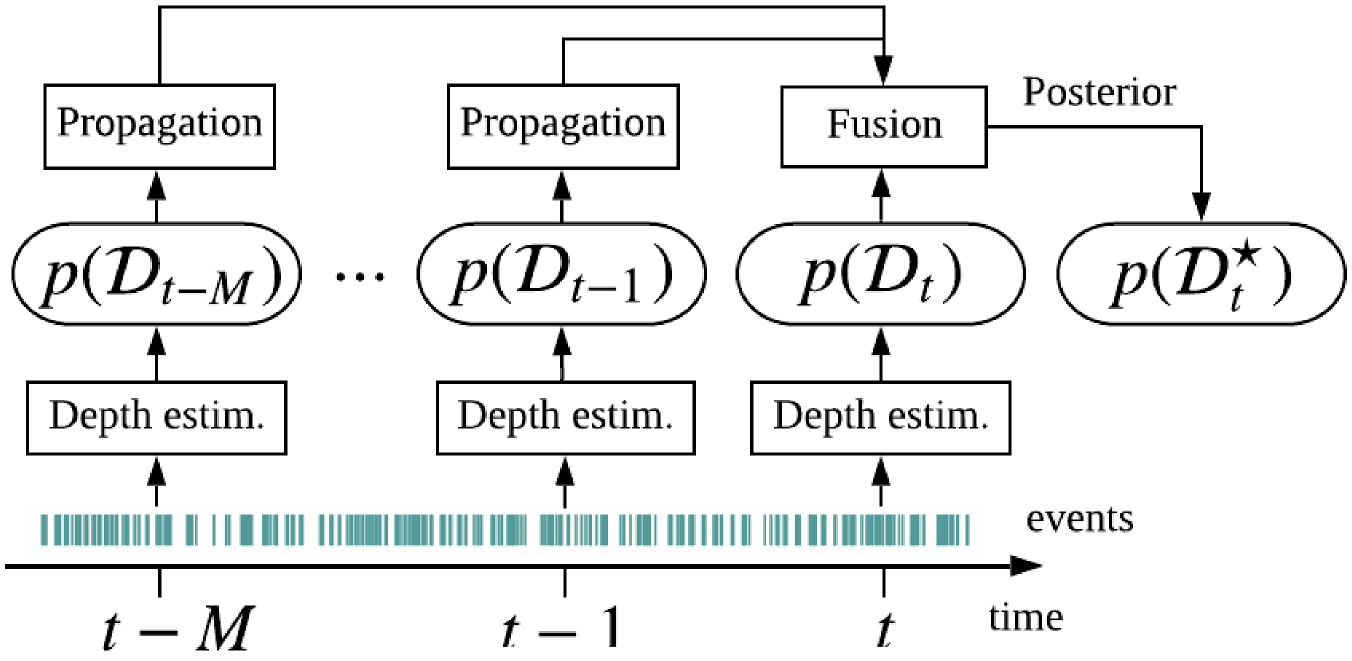


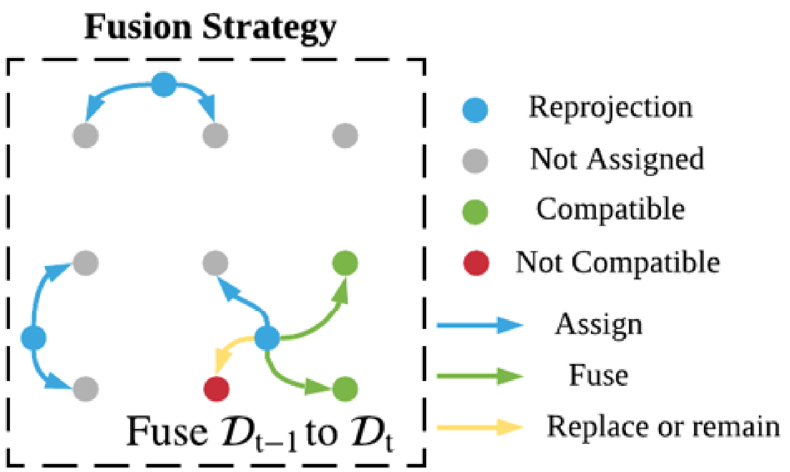
Objective function

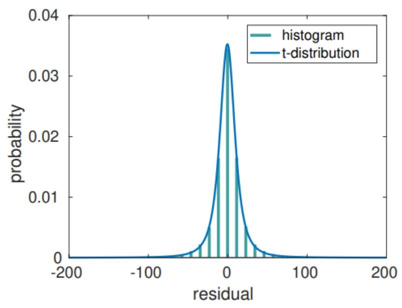
$$\rho^\star = \arg\min_{\rho} C(\mathbf{x}, \rho, \mathcal{T}_{\text{left}}(\cdot,t), \mathcal{T}_{\text{right}}(\cdot,t), \mathbf{T}_{t-\delta t:t})$$

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

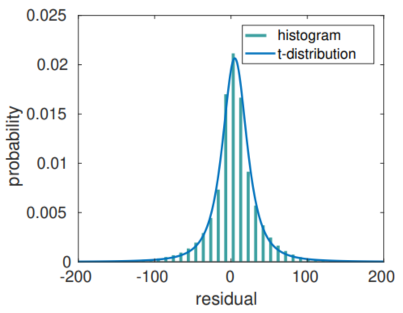
Probabilistic Fusion







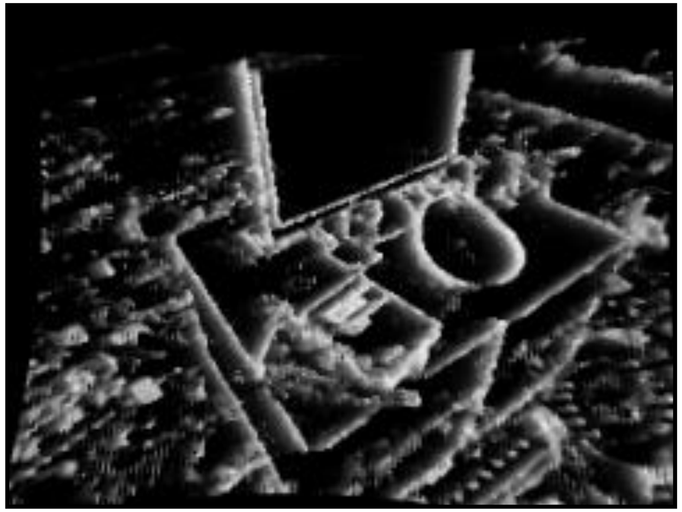
(a) *simulation_3planes* [48].



(b) *upenn_indoor_flying1* [44].

Tracking

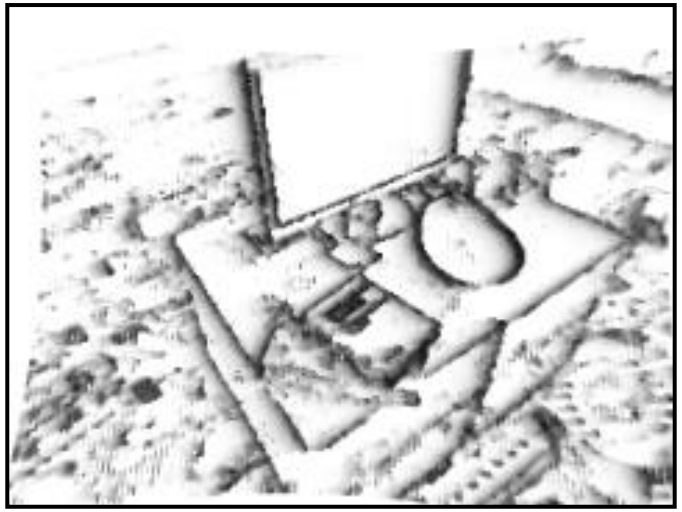
Exploiting Time Surfaces as Distance Fields



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

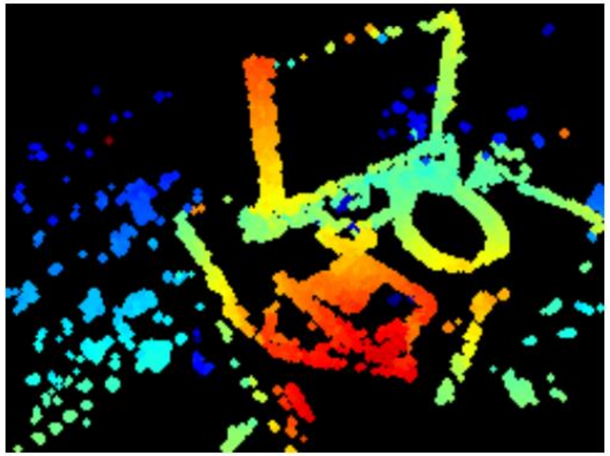
Anisotropic Distance Field

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

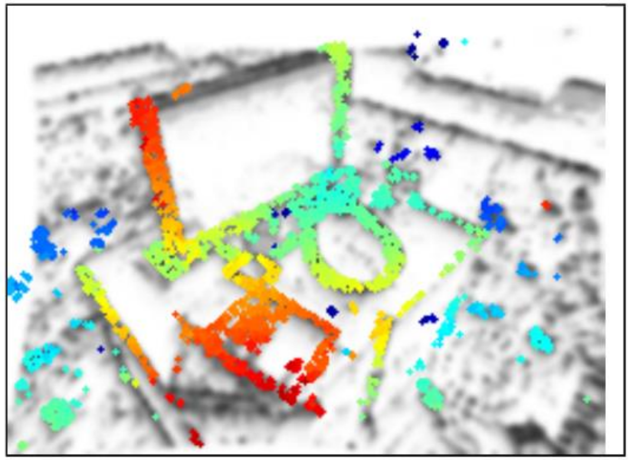


Time Surface Negative $\bar{\boldsymbol{\tau}}(\mathbf{x})$

3D-2D Registration



(a) Inverse depth map in the reference frame.



(b) Registration on the negative time-surface map.

Objective Function

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

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

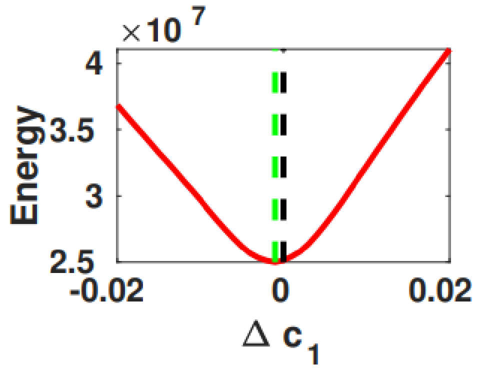
$$\boldsymbol{\theta} \doteq [c_1, c_2, c_3, t_x, t_y, t_z]^T, \quad G(\boldsymbol{\theta}) : \mathbb{R}^6 \rightarrow \mathrm{SE}(3)$$

Forward Compositional LK Method

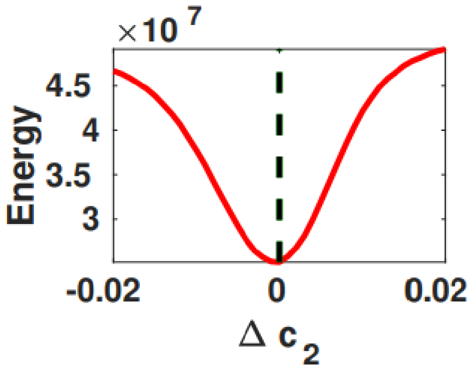
$$F(\Delta\boldsymbol{\theta}) \doteq \sum_{\mathbf{x} \in \mathcal{D}^{\mathcal{F}_{\text{ref}}}} \| \underbrace{\bar{\boldsymbol{\tau}}_{\text{left}}^{\mathcal{F}_k}(W(W(\mathbf{x}, \rho; \Delta\boldsymbol{\theta}); \boldsymbol{\theta}))}_{r_{\mathbf{x}}} \|^2$$

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

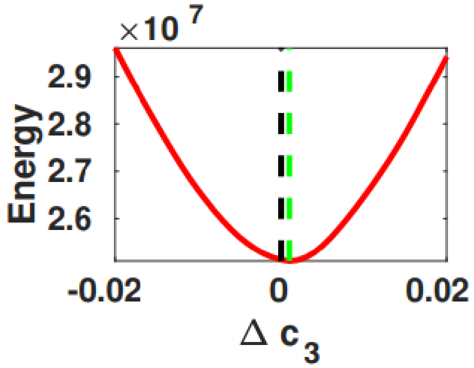
Objective Function



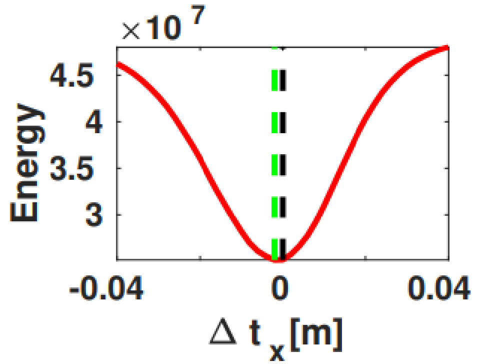
(a) Objective w.r.t c_1.



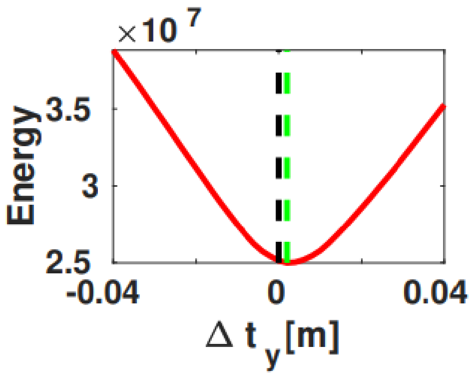
(b) Objective w.r.t c_2.



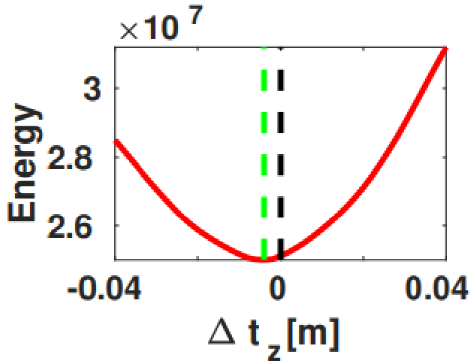
(c) Objective w.r.t c_3.



(d) Objective w.r.t t_x.



(e) Objective w.r.t t_y.



(f) Objective w.r.t t_z.

Evaluation

TABLE II: Parameters of various stereo event-camera rigs used in the experiments.

Dataset	Cameras	Resolution (pix)	Baseline (cm)	FOV (°)
Zhou *et al.*(ECCV2018)	DAVIS240C	240×180	14.7	62.9
Zhu *et al.*(RAL2018)	DAVIS346	346×260	10.0	74.8
Mueggler *et al.*(IJRR2017)	Simulator	346×260	10.7	74.0
Ours	DAVIS346	346×260	7.5	66.5

Our stereo event camera rig set up.

Conclusion

Summary

1. Provide a brief literature review on event-based VO and point out the core problem in the design.
2. Disclose technical details of our recent work – Event-based Stereo Visual Odometry.

Project 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 Star 202

Take-Home Messages

1. Trade-off between latency and computation complexity.

   Bash v.s. Event-by-Event

2. Computational resource and power consumption.

Goal: compact and energy-efficient solution.

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