Structural Information Leakage in Event-Based Camera Streams Without Explicit Reconstruction

Research question: What types of structural scene information can be inferred from raw event-based data without explicit image reconstruction?

Author: Ilinca Mocanu
Responsible Professor: Nergis Tömen
Supervisor: Tunahan Parlayici

01 Introduction

Event cameras

- Event cameras record asynchronous brightness changes per pixel as **(x, y, t, p)** — not full images.
- Useful for fast-motion settings such as camera-surveillance and autonomous driving.

Privacy assumption

- Since event cameras do not directly capture colour, texture, or full image frames, they are often seen as more **privacy-preserving** than RGB cameras.
- However, the absence of normal images does not mean that information cannot be inferred.

Research gap

- Event-to-video reconstruction shows that event streams can be converted into human-readable intensity video [1]. Related privacy work also studies event-camera visual localization and risks around event-to-image reconstruction [2].
- Less is known about what structural information can be inferred directly from **event representations before** any image or video is reconstructed.

02 Objective

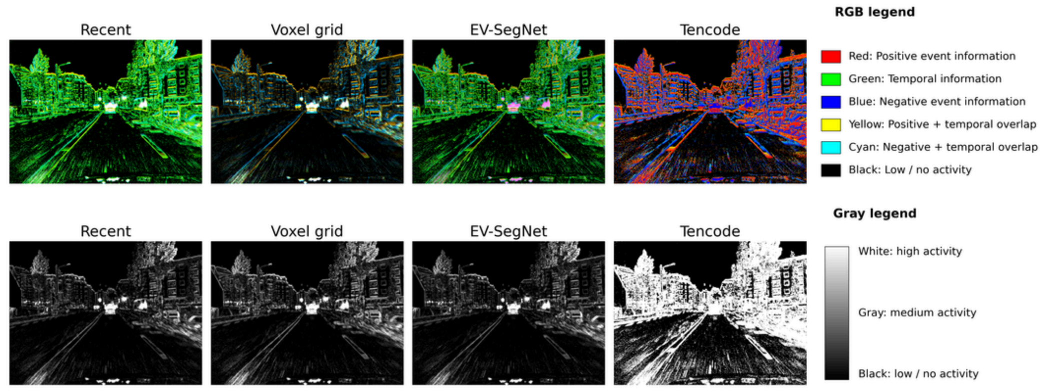
Structural information

- Structural information = visual information about how a scene is organized.
- This includes object contours, human location, human silhouettes, spatial depth structure, and motion patterns.

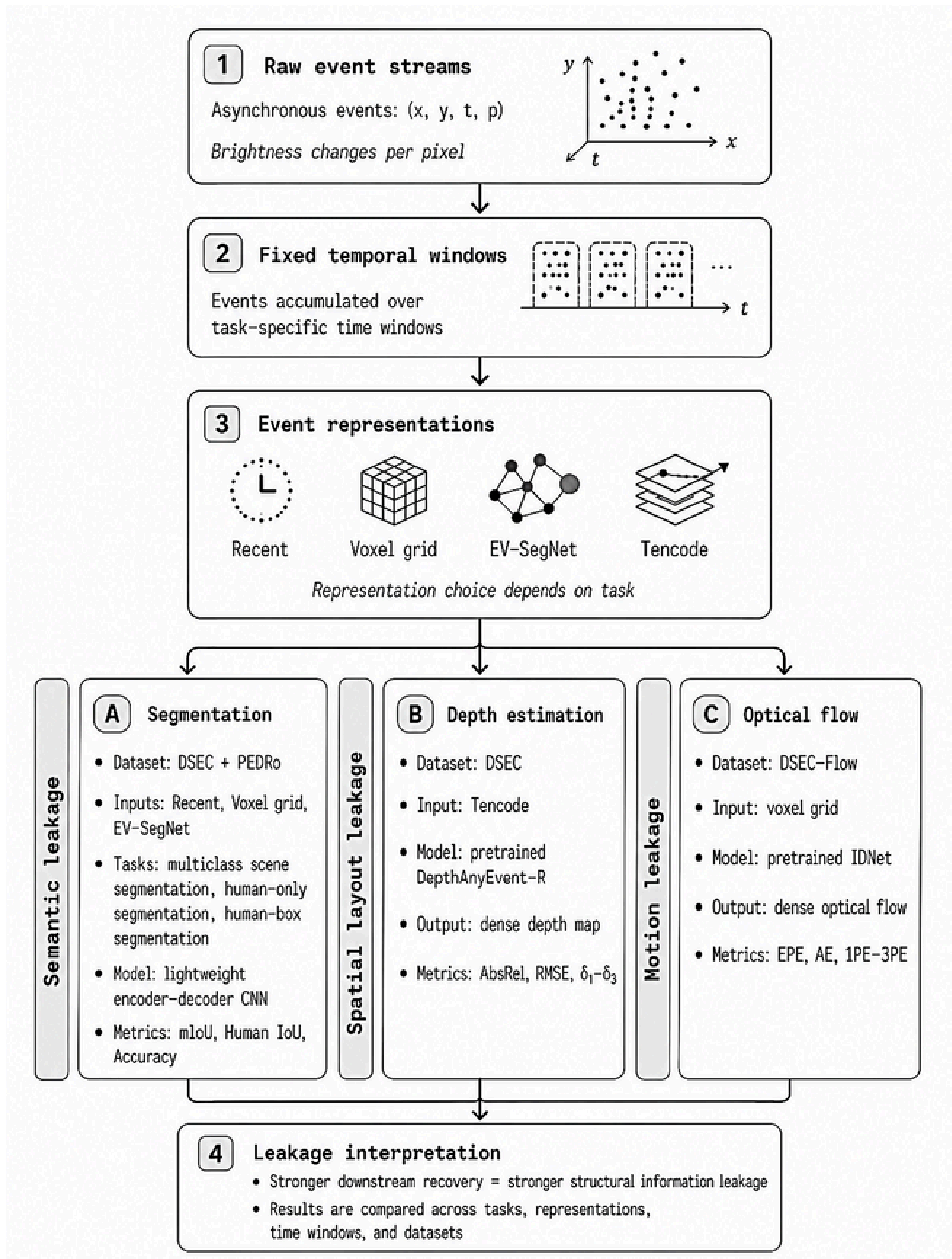
To address the research gap, the project evaluates three forms of leakage:

- **Semantic leakage** → can object classes and human regions be recovered through segmentation?
- **Spatial layout leakage** → can scene depth and distance structure be estimated?
- **Motion leakage** → can dense motion patterns be recovered through optical flow?

If a model can recover structure from an event representation, then that representation exposes privacy-relevant information.



03 Methodology



04 Results

DSEC segmentation results. Values are reported as mean ± standard deviation over three seeds.

Representation	Time window	mIoU	Human IoU
Recent	10 ms	0.385 ± 0.023	0.009 ± 0.015
Recent	50 ms	0.407 ± 0.007	0.007 ± 0.013
Recent	250 ms	0.354 ± 0.007	0.000 ± 0.000
Voxel grid	10 ms	0.386 ± 0.013	0.027 ± 0.034
Voxel grid	50 ms	0.410 ± 0.010	0.023 ± 0.028
Voxel grid	250 ms	0.362 ± 0.010	0.024 ± 0.032
EVSegNet	10 ms	0.398 ± 0.013	0.009 ± 0.008
EVSegNet	50 ms	0.415 ± 0.002	0.016 ± 0.015
EVSegNet	250 ms	0.353 ± 0.012	0.002 ± 0.001

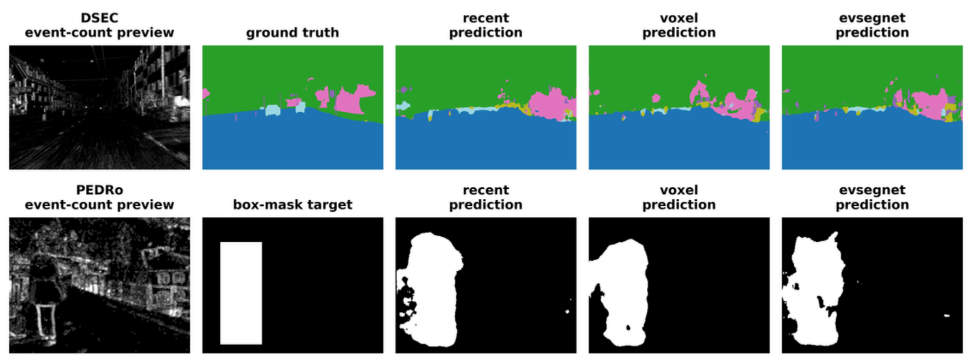
PEDRo human-box segmentation. Bars show mean values over three seeds, with error bars indicating standard deviation.

Depth estimation. Average DepthAnyEvent-R results across evaluated DSEC sequences using the Tencode event representation. Lower is better for AbsRel/RMSE; higher is better for δ metrics.

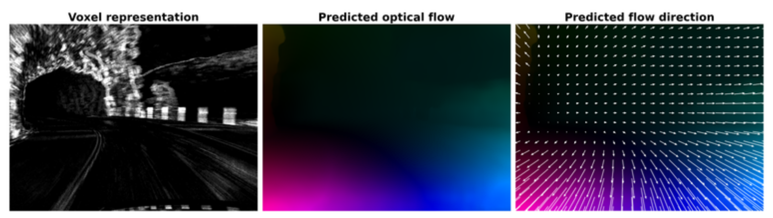
AbsRel ↓	RMSE ↓	δ1 ↑	δ2 ↑	δ3 ↑
0.203	8.853 m	0.665	0.916	0.975

Optical flow estimation: Average IDNet results across evaluated DSEC sequences using the voxel-grid event representation. Lower values indicate better motion estimation.

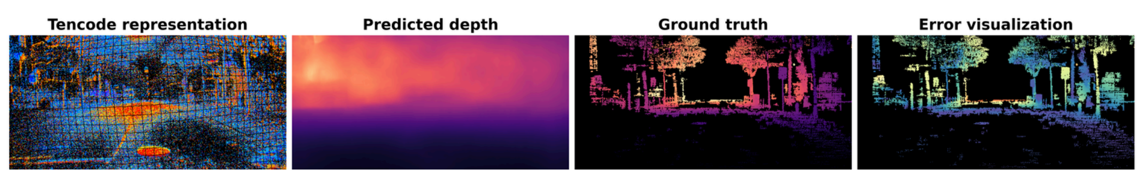
EPE ↓	AE ↓	1PE ↓	2PE ↓	3PE ↓
0.771	3	12.102	4.028	2.274



Segmentation example: The DSEC example shows recovery of large semantic scene regions, while the PEDRo example shows approximate human-box localization from event representations.



Optical-flow example: Qualitative optical-flow result using IDNet. The voxel representation is used to predict dense motion, with the final panel showing sampled flow directions over the predicted flow field.



Depth example: Qualitative depth estimation result using DepthAnyEvent-R. The Tencode event representation is converted into a dense depth prediction and compared with DSEC ground truth and the corresponding error map.

05 Conclusion

- Event representations preserve several types of structural information even without explicit image reconstruction.
- Semantic leakage is uneven. DSEC mainly shows partial recovery of scene layout, better for large classes such as roads or sky, while human IoU remains low because of its sparity in the dataset. PEDRo gives clearer evidence of human-location leakage, but only as filled bounding-box masks, not precise silhouettes.
- Depth and optical-flow results show stronger recovery of non-semantic structure, but both use pretrained task-specific models. These results should therefore be read as strong-model leakage probes rather than a complete comparison of all possible event representations.
- Privacy risk depends on the full pipeline: representation, temporal window, model, task, and dataset all affect what information can be inferred.

References: [1] Henri Rebecq et al. "Events-to-video: Bringing modern computer vision to event cameras". In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019, pp. 3857–3866. doi: 10.1109/CVPR.2019.00398
[2]Junho Kim et al. "Privacy-preserving visual localization with event cameras". In: IEEE Transactions on Image Processing 34 (2025), pp. 6215–6230. doi: 10.1109/TIP.2025.3607640.