EGR 598: Connected and Automated Vehicles

Supporting the Intersection Safety Project team

Project Sponsor: Dr. Yezhou Yang and Dr. Bharatesh Chakravarthi, ASU
Project Engineer: Jayaram Atluri

Arizona State University

Contents

Project Objectives

The objective of this project is to create an intersection safety system that uses AI to combine data from cameras and event sensors. The system will detect vehicles, pedestrians, and other traffic participants to enhance safety and address unsafe conditions. To achieve this, we will set up event cameras and process asynchronous data from the event stream. We will also synchronize and process data to ensure consistency across event data and camera images, which will then aid in the creation of data fusion models for detection. Our objective is to align with the Intersection Safety Project/Challenge by detecting traffic participant movements and improving safety at intersections.

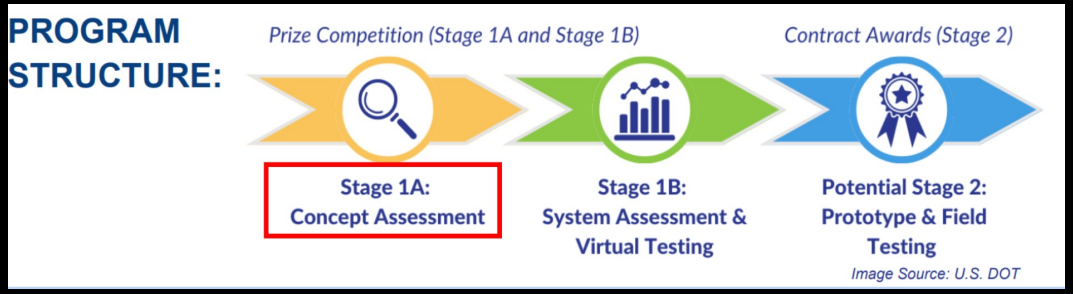
In this presentation, I will discuss the challenges we faced, the current state of the project, the progress we have made so far, and our plans for the future.

INTERSECTION SAFETY CHALLENGE

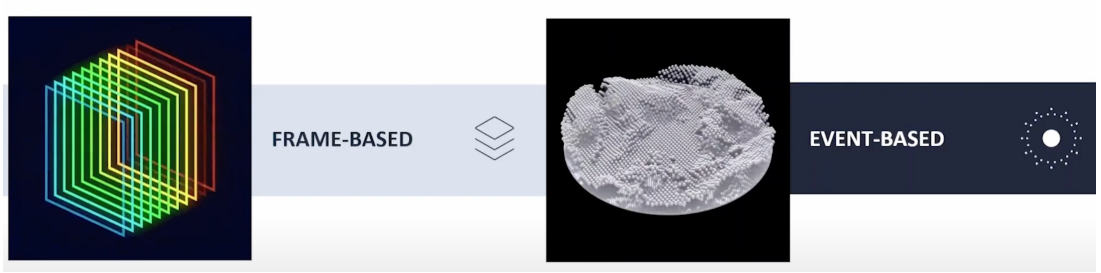
Concept Illustration: Intersection Safety System

Safety systems informed by data fused from multiple sensors may anticipate unsafe conditions, e.g., a vehicle turning right in potential conflict with pedestrian pushing a stroller.

Image Source: U.S. DOT.



Understanding RGB and Event Cameras(Use cases, Setup and calibration) :



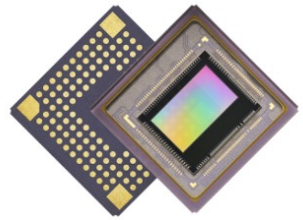
- GENERATES SEQUENTIAL FRAMES
- CLOCK-DRIVEN (PRE-DEFINED FRAME RATE)

- GENERATES CONTINUOUS EVENTS(ASYNCHRONOUS INTELLIGENT PIXELS)
- SCENE-DRIVEN (14s TIMESTAMP RESOLUTION)

Introduction(Keypoints):

1. Event cameras detect changes based on an adjusted threshold (**ON - OFF - Ref**), so events are generated based on our set **contrast sensitivity threshold**.
2. **Sensitivity bias** can be adjusted based on what intensity change should be detected for the application, so data is only generated when crossing the contrast sensitivity threshold.
3. It is observed that **contrast sensitivity** varies with lighting and luminance levels.
4. Higher sensitivity would generate a higher amount of data.
5. So there must be a trade-off between detectable sensitivity and data generation.
6. **Background rate** refers to the number of events generated over time under constant and static illumination, expressed in Hz/px. This results from both photonic noise and electronic noise of the pixel.
7. The **pixel response time** can vary due to changes in contrast and other factors.
8. The **probability density function** (PDF) can be computed by measuring the pixel's response time to a stimulus (e.g. LED on/off) multiple times and collecting statistics.

Industry's Smallest[*1] 4.86µm Pixel Size for Detecting Subject Changes

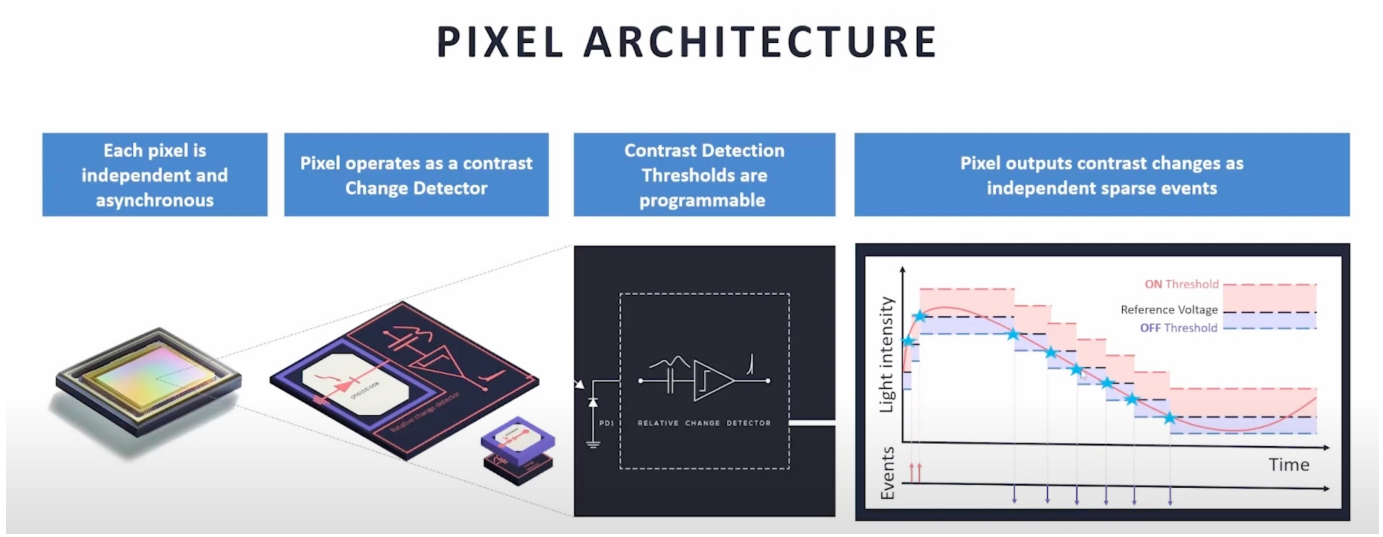


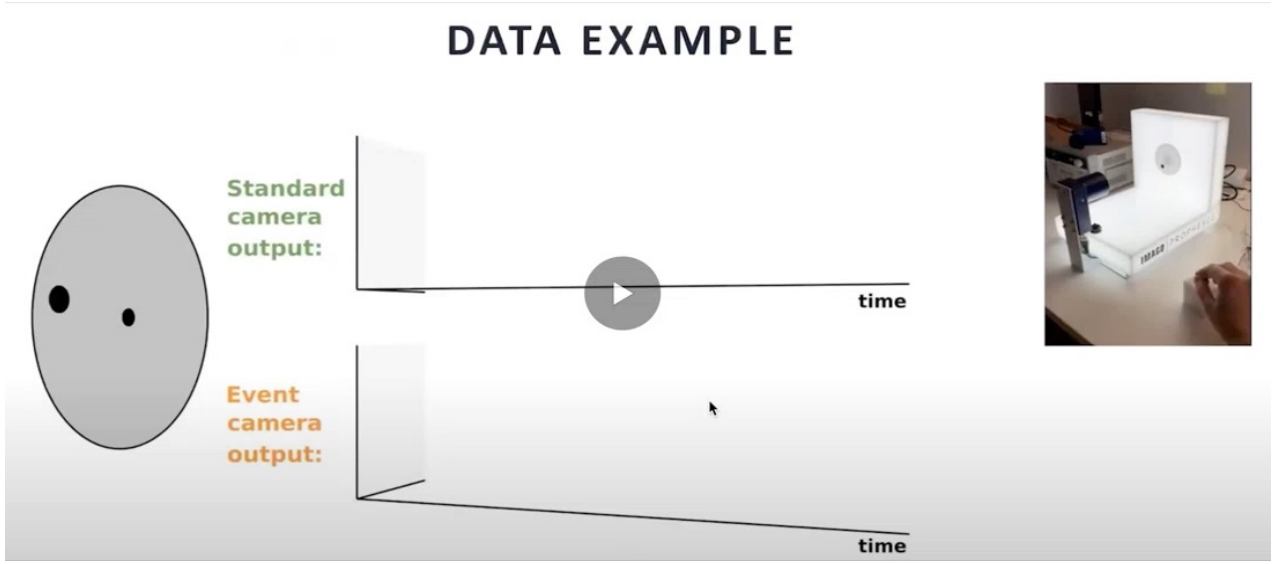
PROPHESEE | SONY

Metavision® sensor IMX636

Understanding RGB and Event Cameras(Use cases, Setup and calibration) :

Continuation….



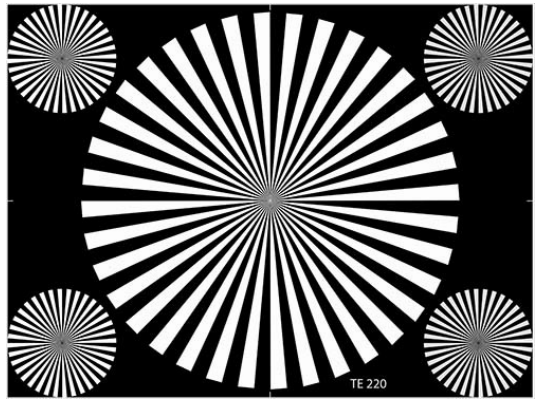


Sources:

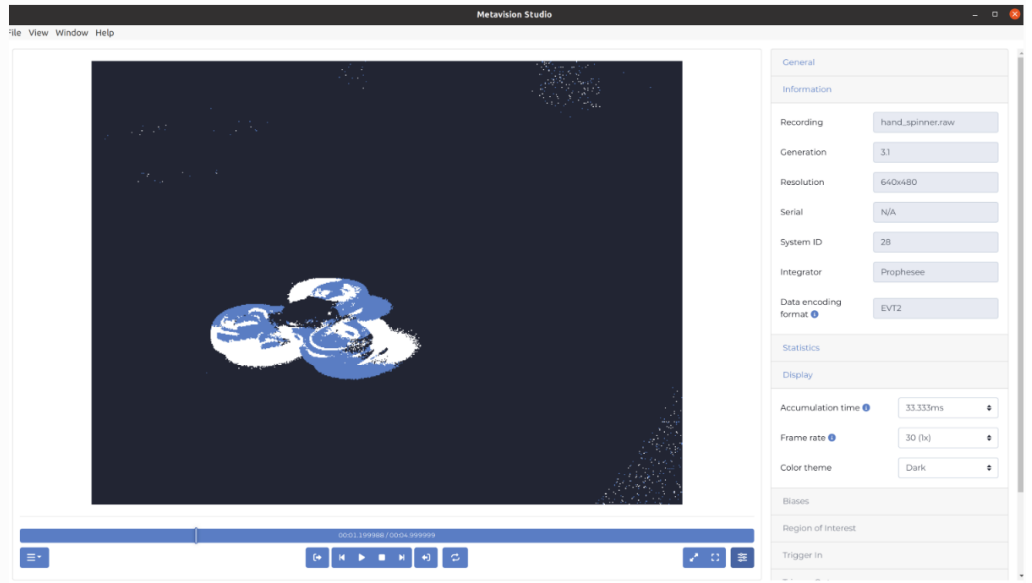
1. https://docs.prophesee.ai/stable/training_videos.html#chapter-tv-bias-tuning

Understanding RGB and Event Cameras(Use cases, Setup and calibration) :

Continuation….



To calibrate event cameras, common methods include using a blinking LED or LCD pattern, an LCD/OLED display, a servo motor rig, or visual-inertial calibration. The most widely used method is the blinking LED board, while the display-based method is cheaper but can introduce errors. Servo rig and visual-inertial calibration are accurate but require complex setups. The image reconstruction method proposed in this paper attached below in sources overcomes previous limitations.



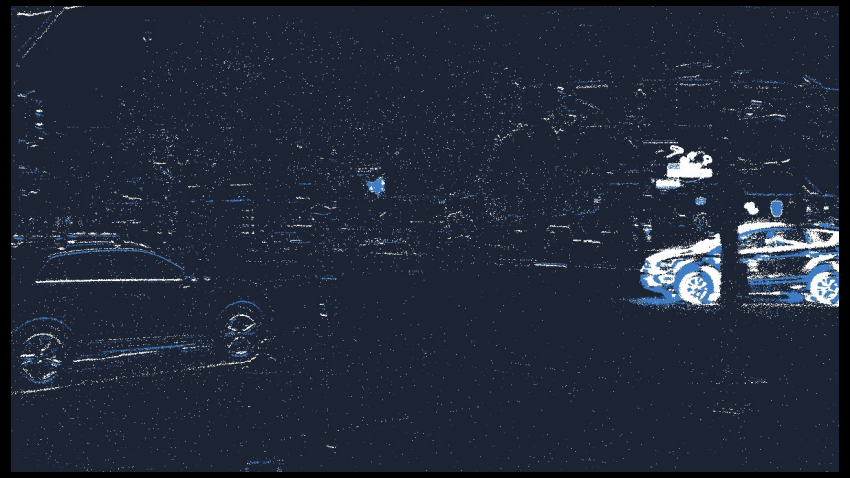
Metavision Studio is an essential tool for working with Prophesee-compatible event-based vision systems. It offers a user-friendly Graphical User Interface that allows us to visualize and record data streamed by these systems. The software comes with RAW tiles in its sample recordings, and when paired with Evaluation Kits or a compatible camera, it makes it easier to adjust the display parameters and tune all the camera settings for optimal performance.

Sources:

1. https://docs.prophesee.ai/stable/metavision_studio/index.html
2. https://tub-rip.github.io/eventvision2021/papers/2021CVPRW_How_to_Calibrate_Your_Event_Camera.pdf

Data gathering And Working on Data synchronization:

RGB camera frame



Event camera frame

Combined result frame for identifying the missing features through individual cameras

I proposed the idea of using this to identify the starting frame which is very crucial for fusion

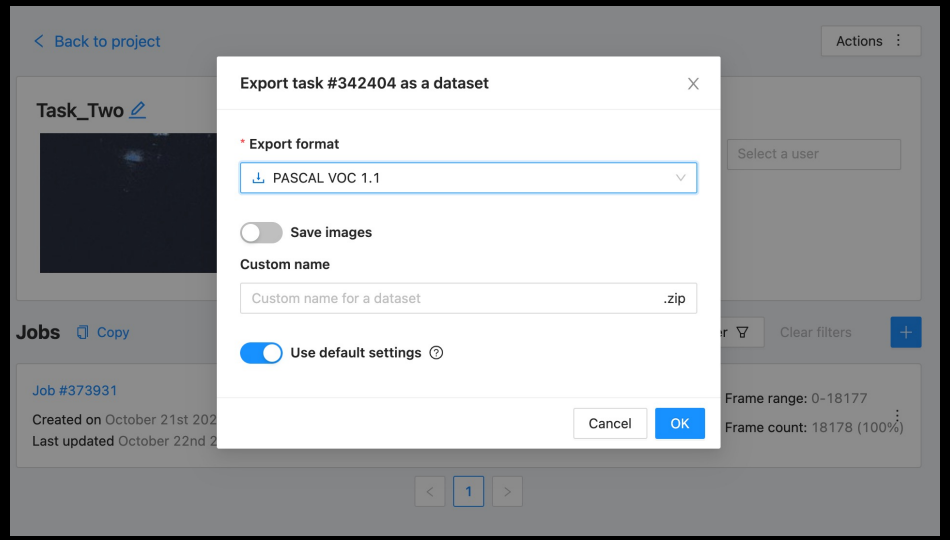
Here each frame is different from others i.e no repetition frames

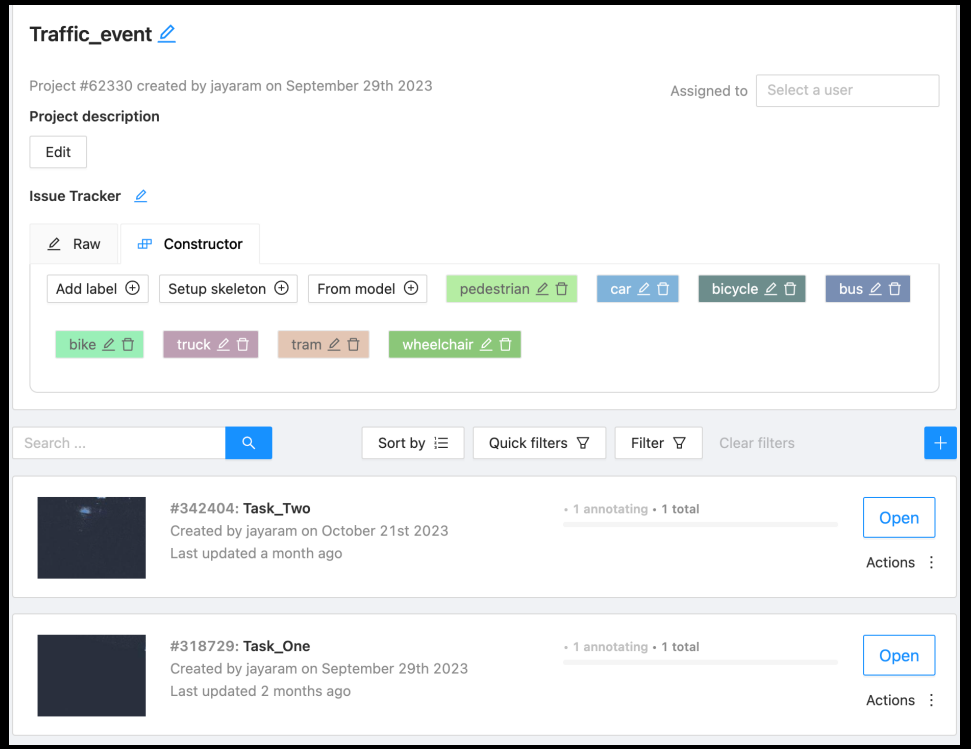
Sources:

1. https://docs.prophesee.ai/stable/hw/manuals/synchronization.html

Sensor Data processing (Labelling, Annotation etc.):

We used CVAT.ai, a computer vision annotation tool, to label multiple object types such as cars, trucks, pedestrians, and bikes using bounding boxes and tracks across frames. The intuitive tools like interpolation between keyframes made labelling long footage feasible. Once the annotation was complete, the labelled dataset was exported into PASCAL VOC 1.1 for ML model training and traffic participant pattern analysis. The cloud access and collaboration features made CVAT.ai optimal for this annotation



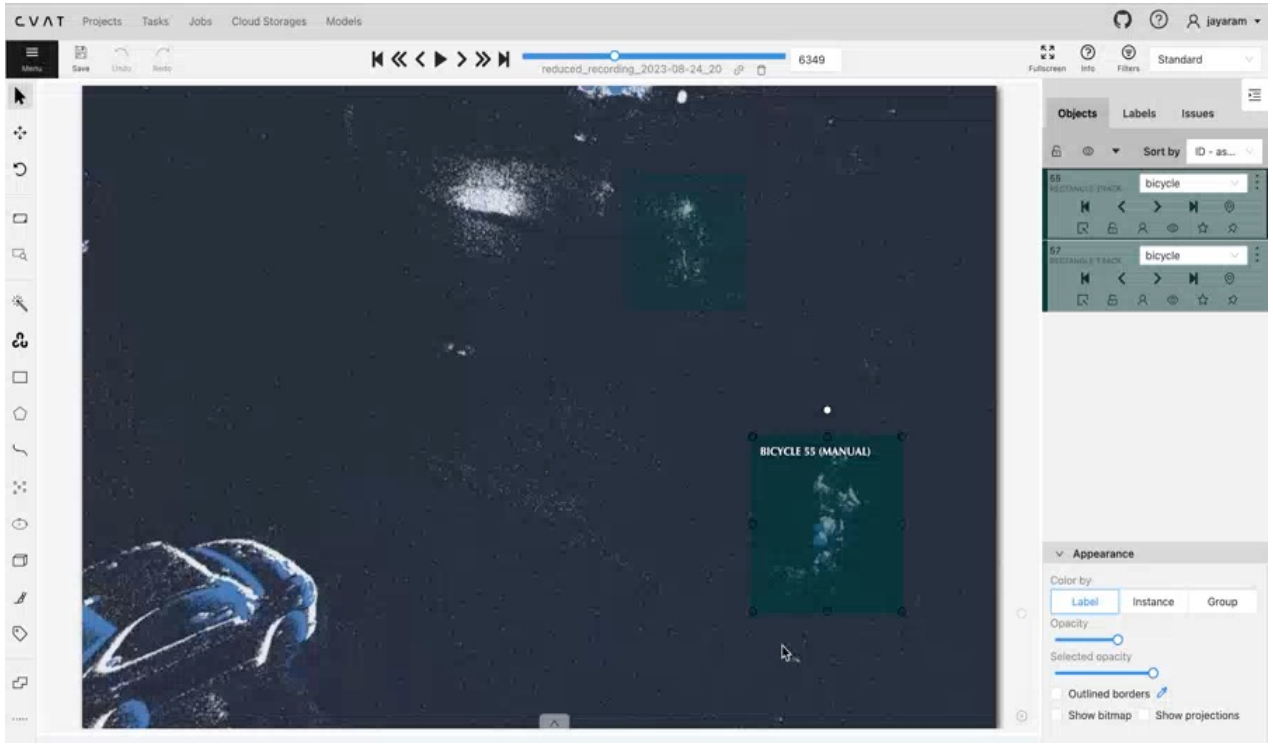


Links:

1. https://www.cvat.ai/
2. https://app.cvat.ai/projects/62330?page=1

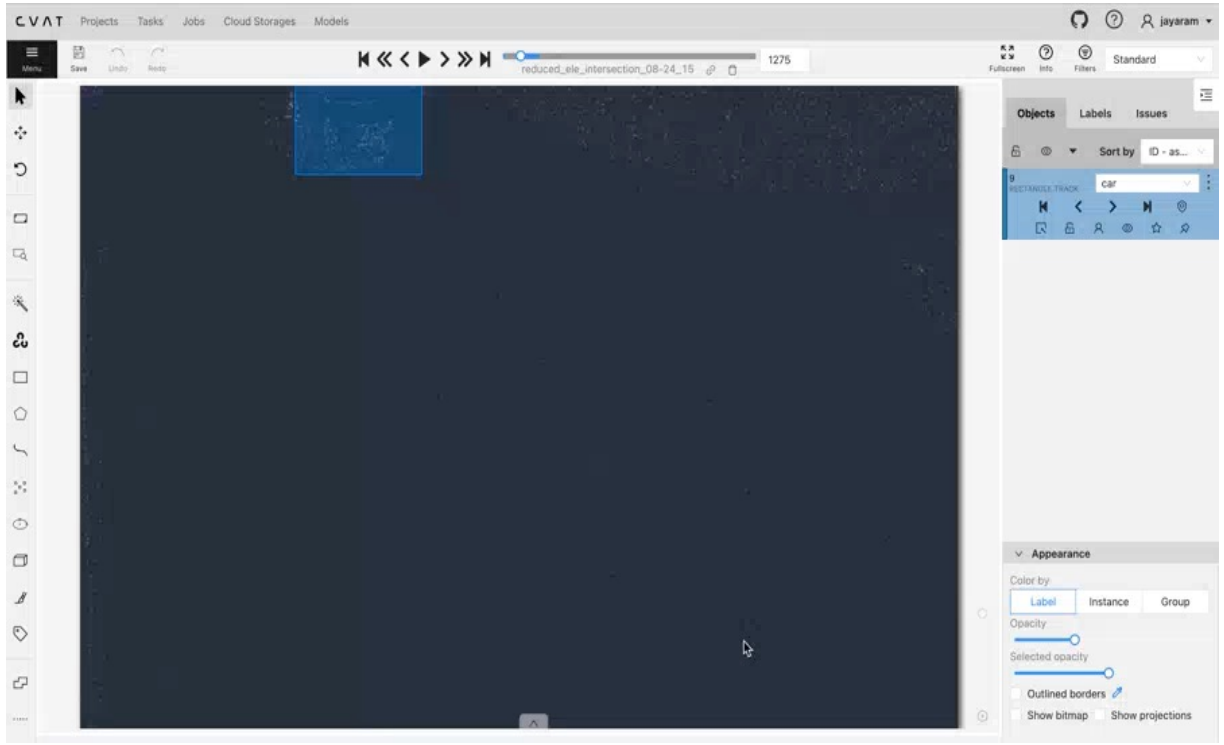
Sensor Data processing (Labelling, Annotation etc.):

Continuation….



Frame count: 18178

https://app.cvat.ai/tasks/342404/jobs/373931



Frame count: 18051

https://app.cvat.ai/tasks/318729/jobs/346195

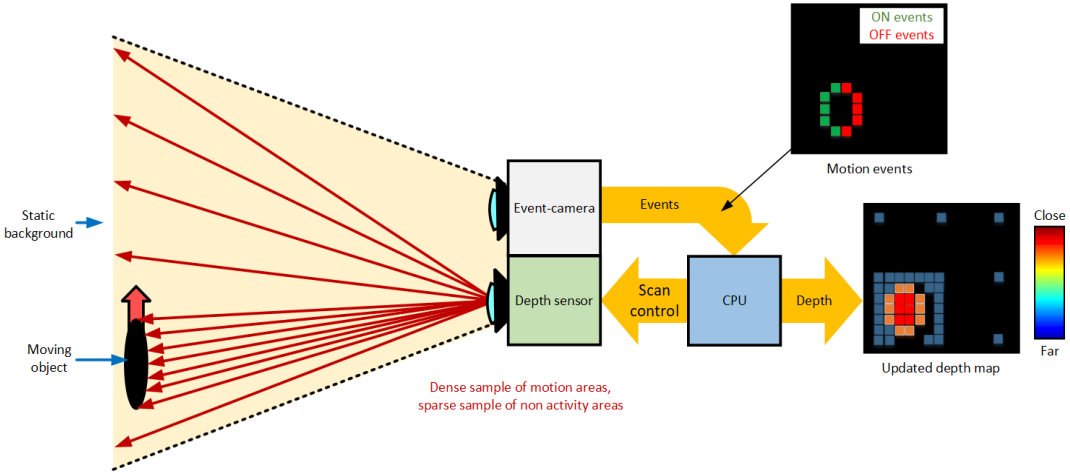
Links:

1. https://app.cvat.ai/tasks/342404/jobs/373931
2. https://app.cvat.ai/tasks/318729/jobs/346195

Sensor Fusion:

Event and RGB sensor fusion requires certain enablers to ensure accurate alignment and timestamping of frames. It is essential to have specific hardware triggers and signals that synchronize the sensors and ensure temporal correspondence between the modalities. Additionally, software alignment techniques such as projection methods and event accumulation over time slices can help map events into the RGB view and register multi-modal data both geometrically and temporally.

In order to fuse event and RGB streams, event projection involves the mathematical projection of oblique event vectors onto a reference RGB image plane, while event accumulation aggregates events into 2D frames over small time intervals, creating event images analogous to RGB frames. Sensor fusion networks can be implemented via early fusion which stacks input streams in multi-channel representations, late fusion which uses deep CNN features concatenated from each stream, or recurrent models which capture temporal event contexts in RNN stacks.



Sources:

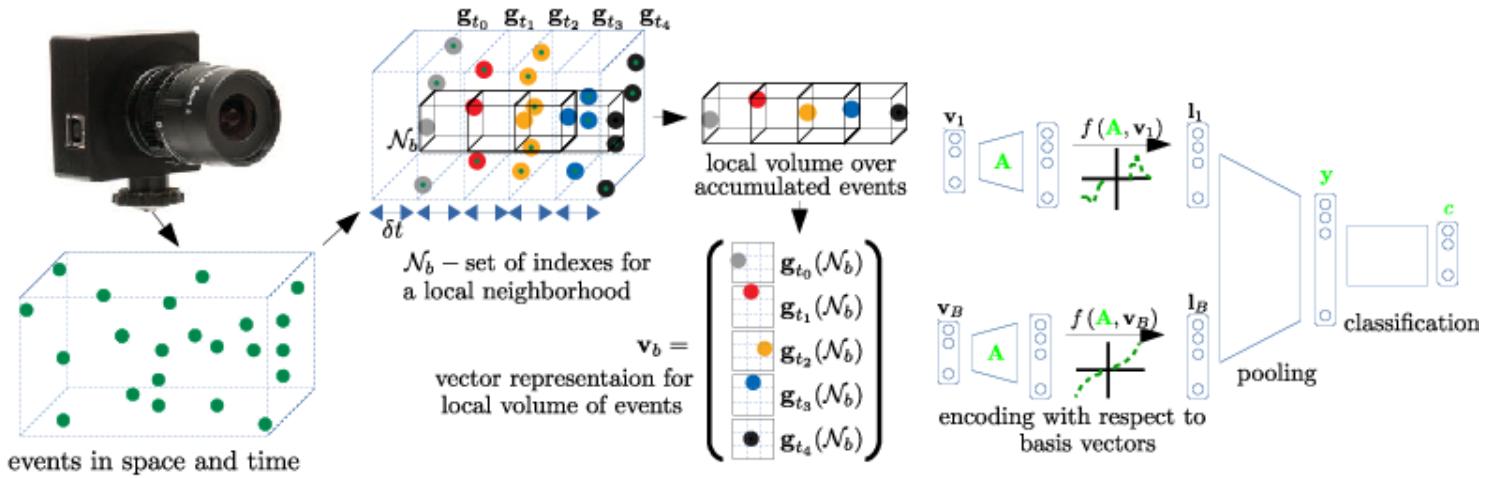
1. https://paperswithcode.com/task/sensor-fusion

Supporting Model Development:

For developing the model, I conducted research on various research papers and articles related to event cameras and models that fuse sensor data from RGB and event cameras for detection and tracking. My findings show that the key approach is to encode the raw event data into spatially organized representations such as event volumes or tensors to enable fusion with RGB frames using standard computer vision techniques.

For object detection, a two-stage model is used that first generates region proposals by combining motion patterns from events with spatial details from RGB images, before classifying the objects. For motion estimation, optical flow leverages the high temporal resolution of events for precise tracking, supplemented by RGB context for handling complex scenes. Trajectory forecasting employs LSTM encoder-decoder networks to learn motion dynamics from event data and improve predictions using RGB scene context.

Finally, anomaly detection classifies abnormal behaviours by analyzing reconstructed event-based images of vehicle trajectories over time, using RGB data to add spatial details for accurate classification. Overall, transforming irregular event streams into grid-like structures enables complex feature fusion and interfacing events with established paradigms requiring regular inputs. This allows the detection, tracking and classifying of various intersection entities for safety.



Sources:

1. https://rpg.ifi.uzh.ch/research_dvs.html#:~:text=Event%2Dbased%20Vision%20meets%20Deep,scene%2C%20filtering%20out%20redundant%20information

Summary and conclusion:

In conclusion, slow-moving objects such as pedestrians and cyclists are better captured through event sensors, making them better suited for our intersection safety situation.

Event sensor cameras are asynchronous and more analog in nature, allowing us to capture all the data from the scene, unlike conventional RGB cameras that capture at fixed FPS rates and may miss data in between frames.

Data is more important in a video when there is a change observed, which is where event cameras are more useful.

Through this project, I have learned that achieving a better perception system is not only about good AI or ML models but also about selecting the right sensors/cameras and fusing them to perceive all events in most real-life situations.

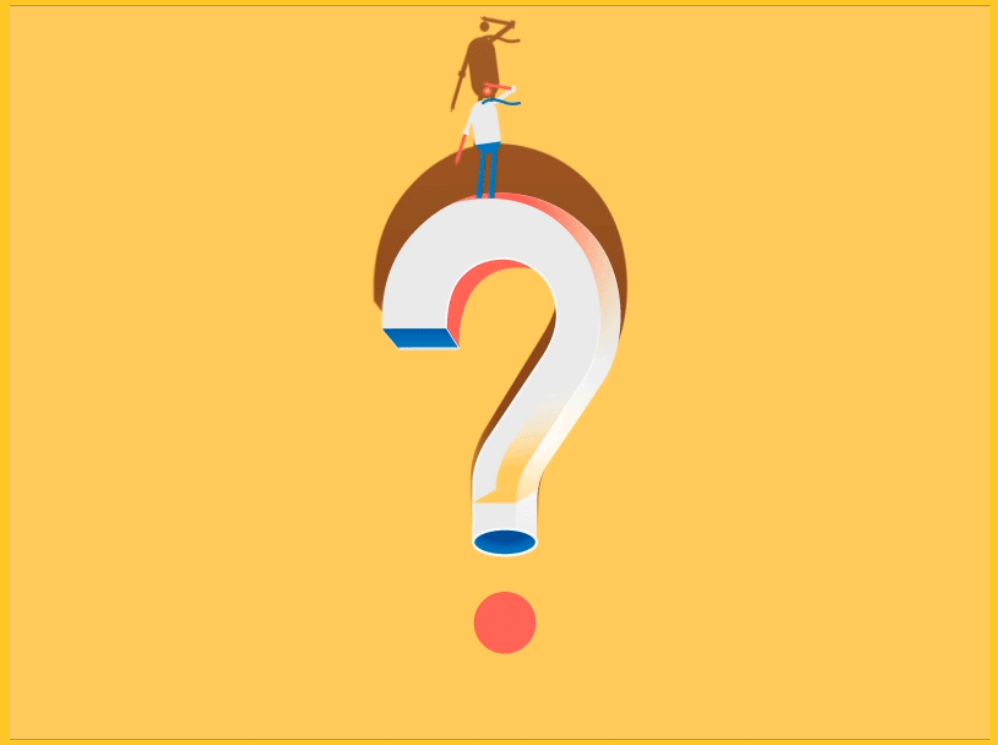
ACKNOWLEDGEMENTS

I am grateful to Dr. Wishart for his invaluable CAV classes. His explanations have helped me to better understand the latest trends and relate them to my ongoing project. I would also like to extend my gratitude to Dr Wishart for giving me the opportunity to work with Dr. Yezhou Yang and Dr. Bharatesh Chakravarthi's team. During this period, I have learned more about cameras and perception. I look forward to continuing my work with the team, particularly in the areas of model development, testing, and deployment.

References:

- Intersection Safety Challenge Website :https://its.dot.gov/isc/
- RFI Summary Report (February 2023): FHWA-JPO-23-986.
- USDOT Intersection challenge submission page: https://www.challenge.gov/?challenge=us-dot-intersection-safety-challenge
- Event-Based Concepts https://docs.prophesee.ai/stable/concepts.html
- https://www.prophesee.ai/event-camera-evk4/
- https://docs.prophesee.ai/stable/hw/manuals/synchronization.html
- https://tub-rip.github.io/eventvision2021/papers/2021CVPRW_How_to_Calibrate_Your_Event_Camera.pdf
- https://paperswithcode.com/task/sensor-fusion
- ASU Standard Presentation Template: https://brandguide.asu.edu/execution-guidelines/presentations
- Image sources: US DOT,CVAT Screenshots,Metavision, Prophesee,Sony sensors
- https://rpg.ifi.uzh.ch/research_dvs.html#:~:text=Event%2Dbased%20Vision%20meets%20Deep,scene%2C%20filtering%20out%20redundant%20information
- https://rpg.ifi.uzh.ch/research_dvs.html

Any Questions




Thank you!