Real-time, Speed-invariant, Vision for Robotics

CVPR2023 Workshop on Event-based Vision

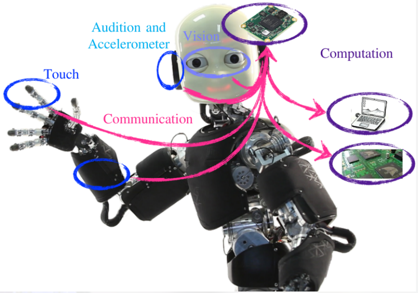
Arren Glover

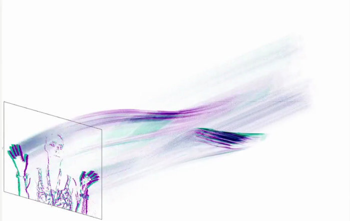
Event-driven Perception for Robotics - Italian Institute of Technology

19th June 2023

Event-driven Perception for Robotics

- Italian Institute of Technology - Home of the "Neuromorphic iCub"
- Vision with event-cameras together with spiking touch sensors and neuromorphic processing
- End Goal: Spiking networks mimicking biology on neuromorphic chips
- Short Term: Demonstrate the advantages of event-cameras for robotics



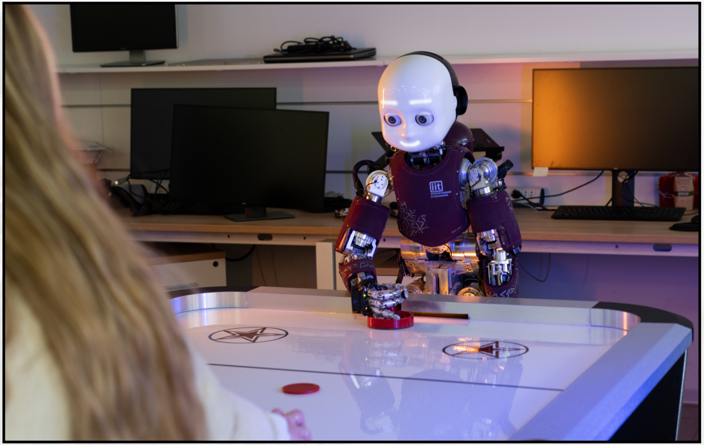


Computer Vision Projects

- Real-time vision algorithms with a live streaming camera in closed-loop robotic control.
- Moving cameras, moving objects and people with unknown and variable speeds.

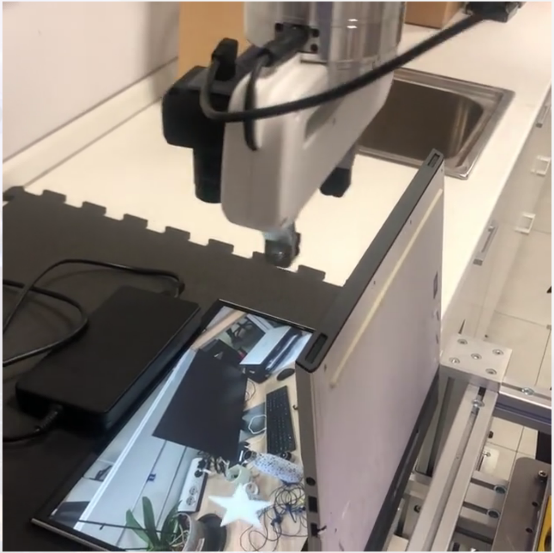
–

- Show our work in:
  - Object Tracking
  - Human Pose Estimation
  - Object Trajectory Prediction
  - Feature Detection
- Common methods we use to overcome challenges with event-cameras



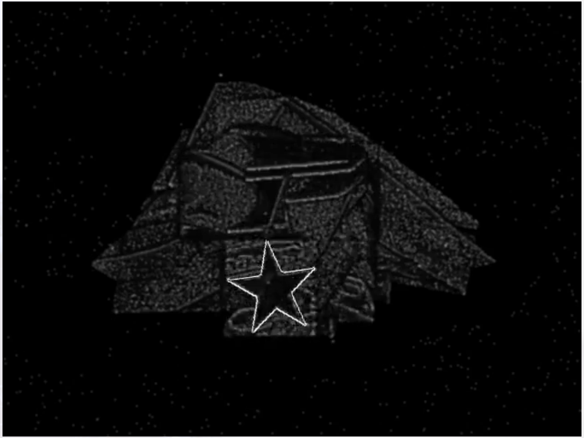
Fast Tracking and Closed-loop Robot Control

- Fast-moving robot in closed-loop control with an event-camera.
- Moving object and moving camera
- Textured background - not a simple clustering exercise
- 4-DoF tracking (x/y position, size, rotation)



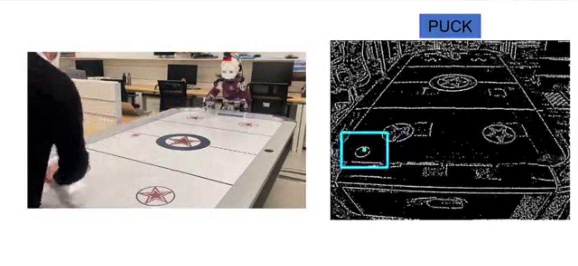
Real-time Affine Shape Tracking

Edges formed from events, tracked star position overlaid.



The robot moves to try to keep the ROI of the star in the centre.

Tracking Applications

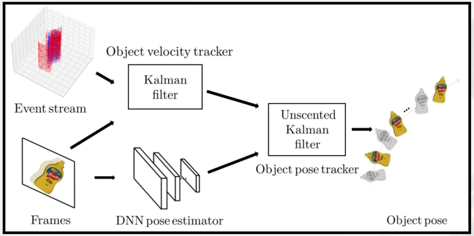


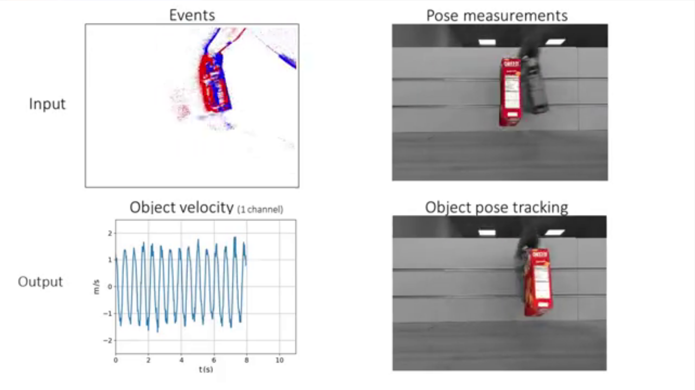
Air Hockey with Event-cameras

Iris tracking - challenging blinks and saccades.

Gava et. al. *PUCK: Parallel Surface and Convolution-kernel Tracking for Event-Based Cameras*. 2022.

Hybrid Frame/Event 6-DoF Tracking



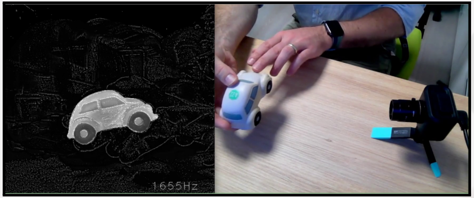


- DOPE 6-DoF pose estimator on frames
- Velocity error from events
- Dual Kalman filters

Events "fill the gaps" of the frame-based detector.

submitted - Li et. al. *Hybrid Object Tracking with Events and Frames*. 2023.

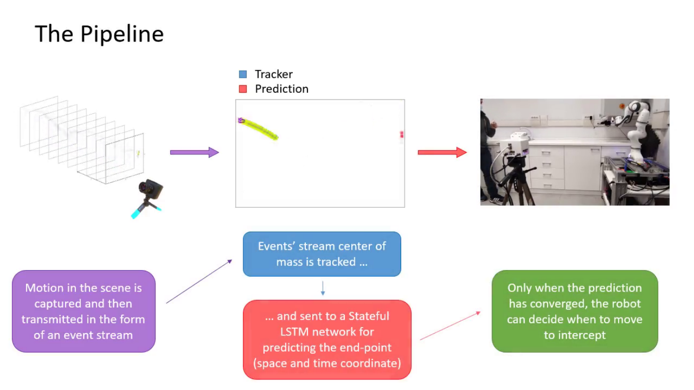
Event-based 6-DoF Tracking



- 3D mesh of a toy car scanned
- 6-DoF pose of car estimated from EROS
- Re-projected onto image-plane.
- Work-in-progress - failure occurs.

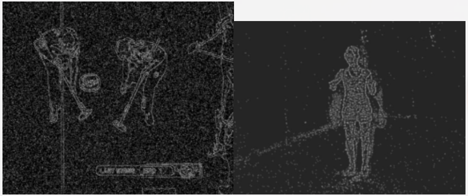
Trajectory Prediction - POSTER

- Detect trajectory, and estimate interception point, within first few pixels
- Robot has more time to arrive at interception location
- Real-sense at 30 Hz - tracking failed due to motion blur.
- Outperformed (earlier prediction and more interceptions) than real-sense at 60 Hz

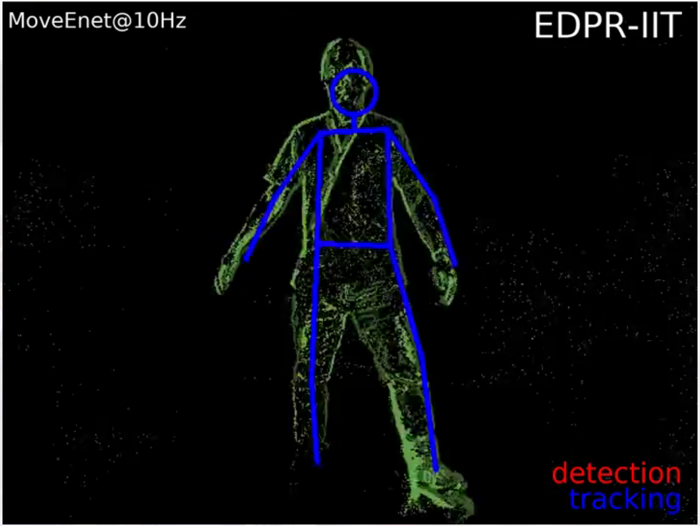


Monforte et. al. *Fast Trajectory End-Point Prediction with Event Cameras for Reactive Robot Control*. 2023.

Human Pose Estimation - POSTER



- "MoveNet" lightweight detection architecture adapted to events
- Exploiting massive frame-based datasets for pre-training with edge-based conversions
- $>$ 1 kHz output when combined with low latency velocity estimation.

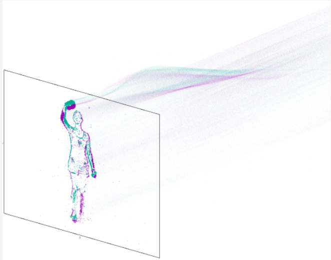


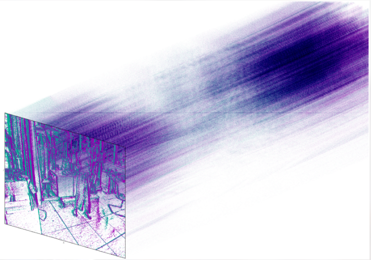
Goyal et. al. *MoveEnet: Online High-Frequency Human Pose Estimation with an Event Camera*. 2023.

Challenges for Robotics

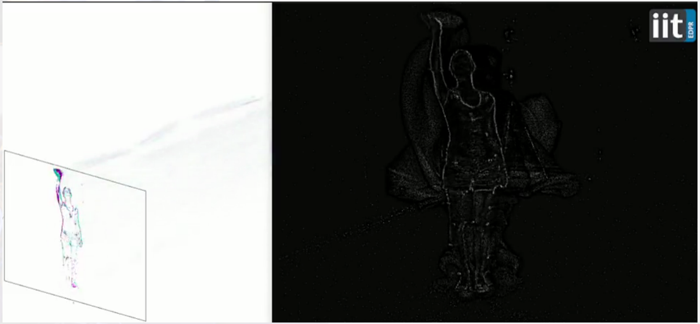
Event-cameras have great potential for robotics, but the advantages also bring about challenges. As event cameras communicate per 1 (or few) events:

1. Event-cameras have a **high temporal resolution**, which results in an **ambiguous temporal association** between pixels.
2. Event-cameras have a **low latency**, but an unknown amount of data in any time-period. **Non-constant processing** rate can break real-time constraints for robots.



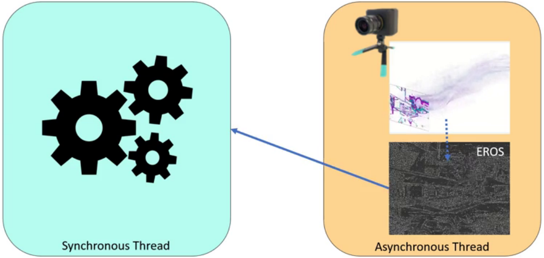


- EROS is a speed invariant surface.
- The element in the 2D array indicates the likelihood of spatial gradient.
- It is updated asynchronously, event-by-event (10M events/s)
- EROS has no temporal decay, and maintains persistence during stopped motion.
- Design decisions leave behind artifacts.



2. Hybrid Synchronous/Asynchronous

- Operations *per-event* are minimised to maximise event throughput.
- **Thread 1**: read events with short processing (update EROS).
- **Thread 2**: complex vision algorithm operates on the data structure in Thread 1 e.g. @1 kHz
- Achieves both a high event-throughput and a complex algorithm running "as-fast-as-possible" (not limited by sensor frequency!).



- Complex algorithms sample the EROS as fast as possible (instead of waiting for *N* events)
- Algorithms are limited by their efficiency, not the sensor FPS

- Speed invariant
- Represents spatial gradients
- Updated asynchronously
- µs resolution

Advantages in Context

Tracking

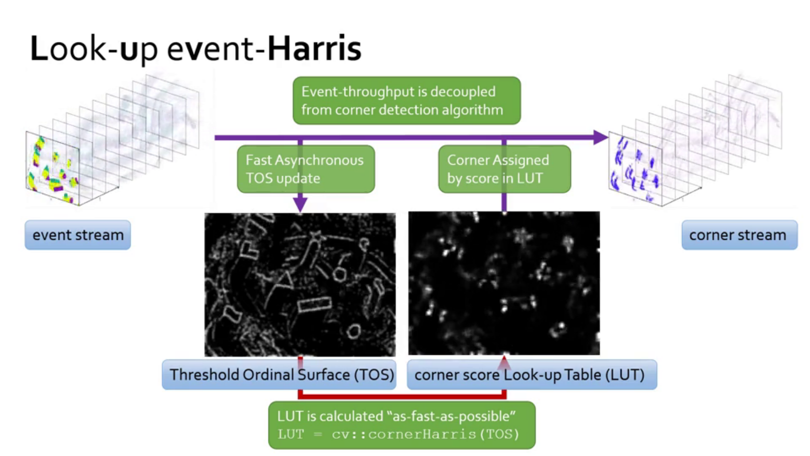
- EROS gives a consistent “appearance” despite variations in speed and size, and moving camera.
- Hybrid processing allows real-time event-throughput (10M events/s) and $>$ 1 kHz tracking rate.

Human Pose Estimation

- EROS gives a consistent “appearance” despite variations limb motion, e.g. waving v.s. stop gesture.
- EROS maintains limb position when stationary, e.g. the leg´s do not disappear.
- EROS unlocked RGB image datasets due to appearance similarity with Canny edge detection.

Asynchronous Corner Tracking with Hybrid Processing

luvHarris maintains an asynchronous output even when using a synchronous thread.



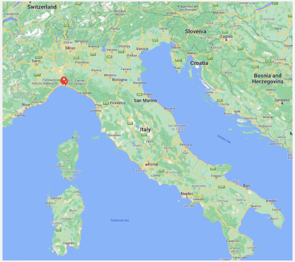
Conclusions

- Design event-camera algorithms under real closed-loop robotics constraints.
- The advantages of event-cameras also brings challenges to be solved.
  - EROS
  - Hybrid Asynchronous/Synchronous

–

- Developing working algorithms can shed light on mechanisms behind biological systems.
- For example - EROS is a method to solve temporal association and persistence; is a similar mechanism found in biology?

Questions



ISTITUTO ITALIANO
DI TECNOLOGIA