Event-based datasets for classification and pose estimation

James P. Turner [1] Jens E. Pedersen [2]
Jörg Conradt [2] Thomas Nowotny [1]

[1] University of Sussex
[2] KTH Royal Institute of Technology

J.P.Turner@sussex.ac.uk

Neuromorphic Vision Data

UNIVERSITY OF SUSSEX

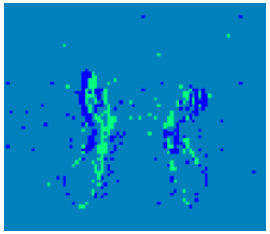
- Many datasets allowed the significant growth of ANN-based deep learning

IMAGENET

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

- Some new spiking event-based datasets for SNN machine learning

IBM DVS128 Gesture Dataset

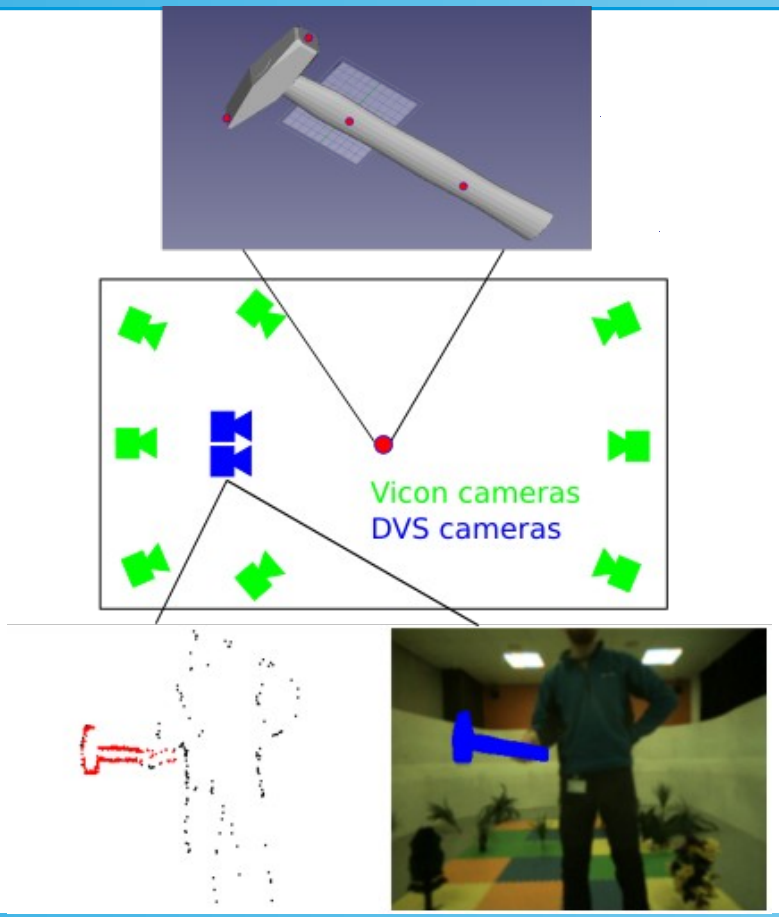


Neuromorphic N-MNIST Dataset

- Issue: lack of infrastructure for collecting large event-based datasets
- Can collect lots of data, but how do we label it? Need automation...

Setup Overview

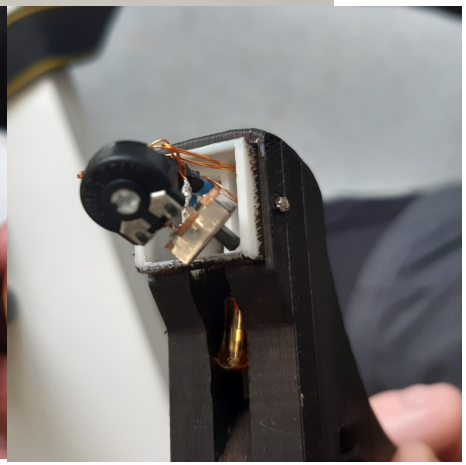
- Goal is to create spiking event-based dataset for human-robot interaction (a robot ‘helper’ which can manipulate tools)
- 8x Vicon Vero 2.2 (3D tracking) cameras
- 2x DVS DAVIS 346 (event/RGB) cameras
- Project prop meshes into Vicon space, based on known marker locations in STL file and Vicon tracker
- Then project props onto DVS camera planes, based on known camera location and orientation



Dataset Props

- Problem: Vicon NIR strobe interferes with DVS cameras, and standard (passive) markers are visible on props
- Solution: custom 3D printed hollow props, with small NIR LED markers and battery

  STL files available online at
  https://github.com/jamesturner246/active-tracker-props
- Remaining low-power NIR light filtered from DVS cameras with 780 nm cut filters



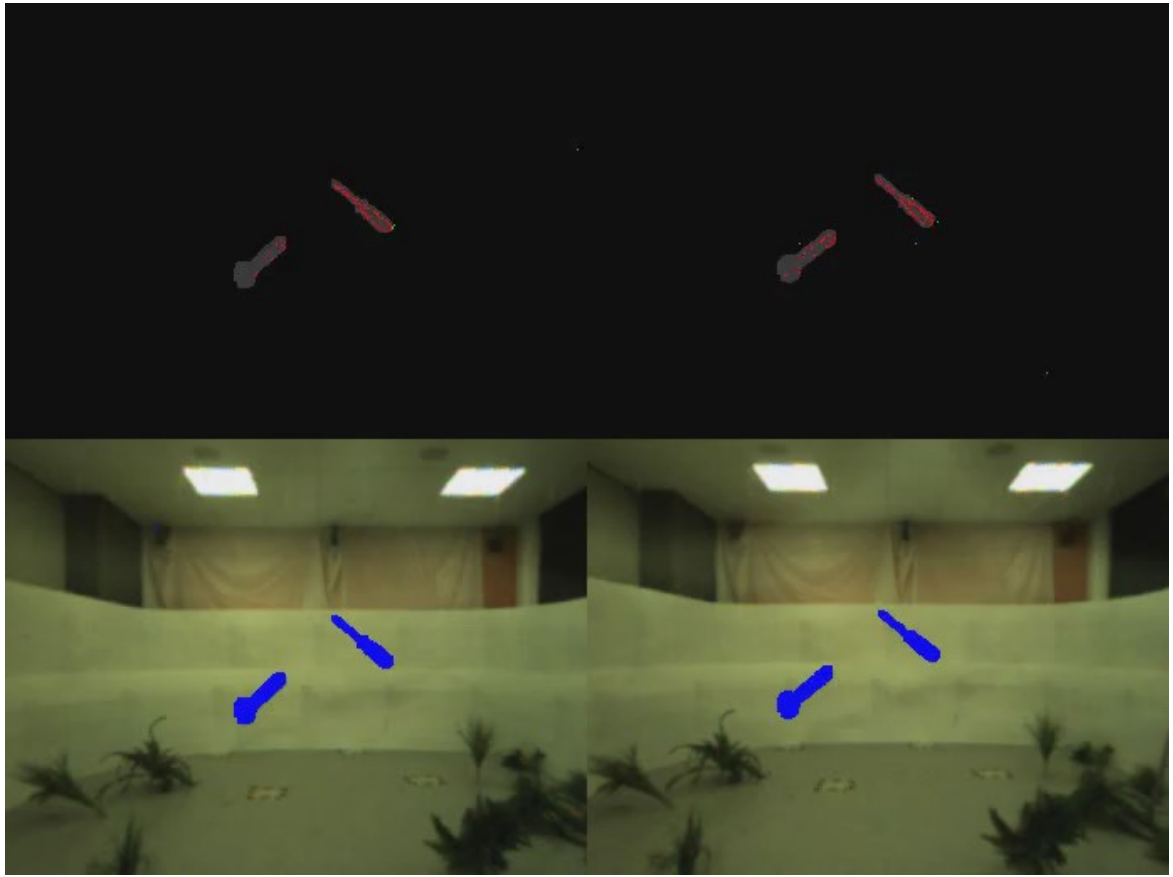
Calibration Procedure

UNIVERSITY OF SUSSEX

- For projection, we need to know the transformation from 3D Vicon space to 2D camera plane coordinates, but camera location and orientation is unknown
- Record flashing markers with Vicon and DVS cameras at several positions (we use the Vicon Active Wand v2, with strobe enabled)
- Through optimisation, we find the best rotation and translation which fits these points from 3D Vicon space to 3D camera-centric space, and project onto the 2D image plane using the standard Pinhole camera model

Example Recording

- 2 cameras, multiple props in view
- Segmentation labels of events and RGB frames
- Translation and rotation (pose) labels
- Extrapolation of bad pose data



Sample Dataset

- Small sample dataset with 9 separate 30 second recordings of suspended moving props
- Stored in HDF5 format online

  https://doi.org/10.25377/sussex.17112080.v1
- Processing code is available online

  https://github.com/jamesturner246/vicon-dvs-projection

./data/[prop]/[sample #]

frame.h5

timestamp_*i*, image_raw_*i*,

image_undistorted_*i*, label_*i*

event.h5

timestamp_*i*, polarity_*i*, xy_raw_*i*

xy_undistorted_*i*, label_*i*

pose.h5

timestamp, extrapolated[*p*]

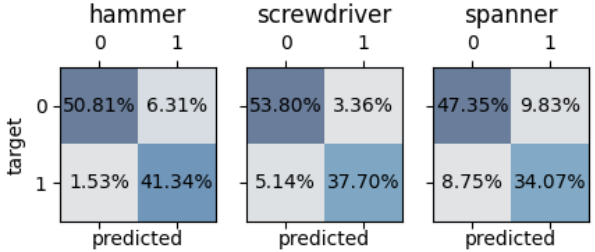
rotation[*p*], camera_rotation_*i*[*p*]

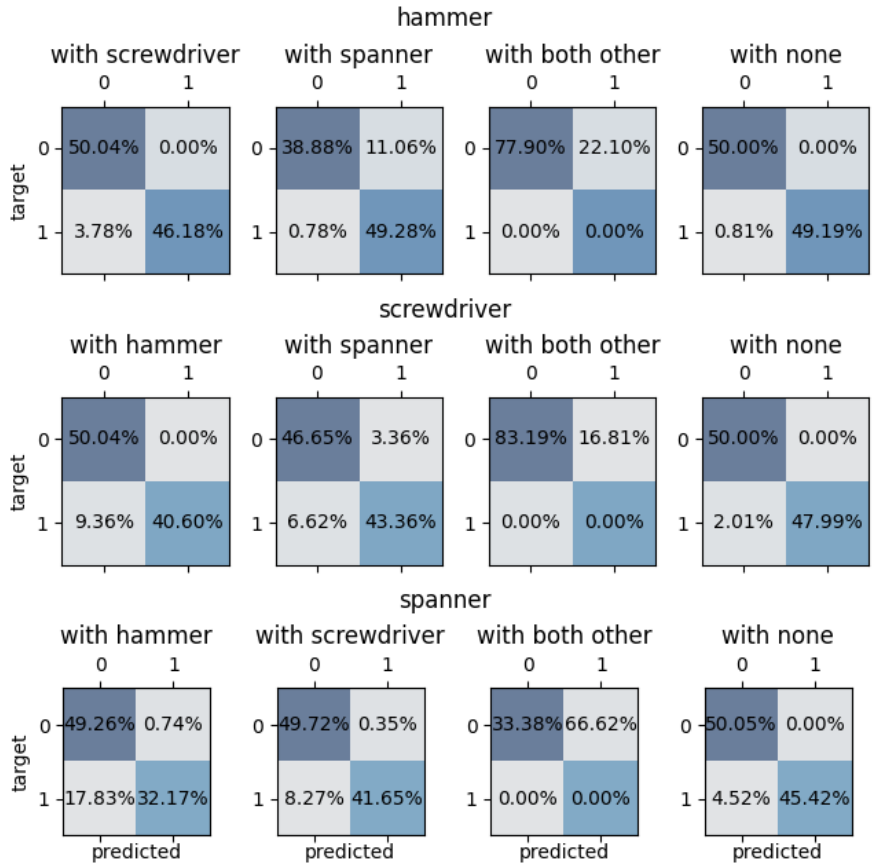
translation[*p*], camera_translation_*i*[*p*]

Event-Based Object Detection

- Events binned into 21msec ‘frames’, with augmentation
- VGG16-like ANN network, trained with transfer learning
- Converted to SNN using Few-Spike Conversion [1]

[1] Stöckl, C., Maass, W. Optimized spiking neurons can classify images with high accuracy through temporal coding with two spikes. Nat Mach Intell 3, 230–238 (2021). https://doi.org/10.1038/s42256-021-00311-4





Issues and Future Work

UNIVERSITY OF SUSSEX

- No events when props are still
- Difficult to disambiguate props at certain angles
- Noise, presence of other objects, occlusions
- Should integrate detection and pose information over time
- Estimate prop translation and rotation estimation, in addition to tool identity

Thanks for your time

Questions?

J.P.Turner@sussex.ac.uk

Human Brain Project