Computer vision application for events cameras

Francisco Javier Martín Ameneiro

franciscojavier.martin@estudiante.uam.es

Universidad Autónoma de Madrid

Escuela Politécnica Superior

Video Processing and Understanding Lab

Mobi Net Video

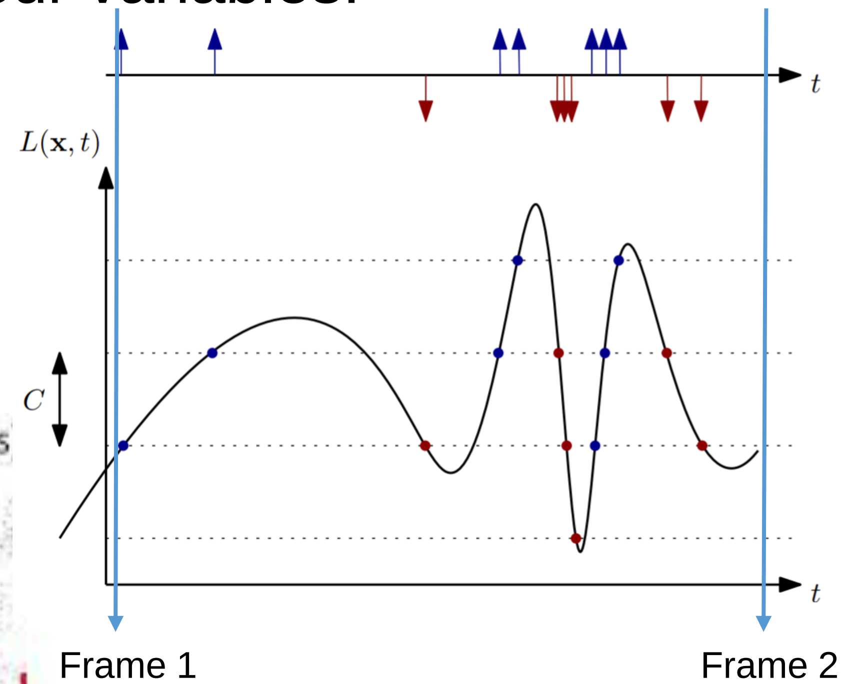
TABLE OF CONTENTS

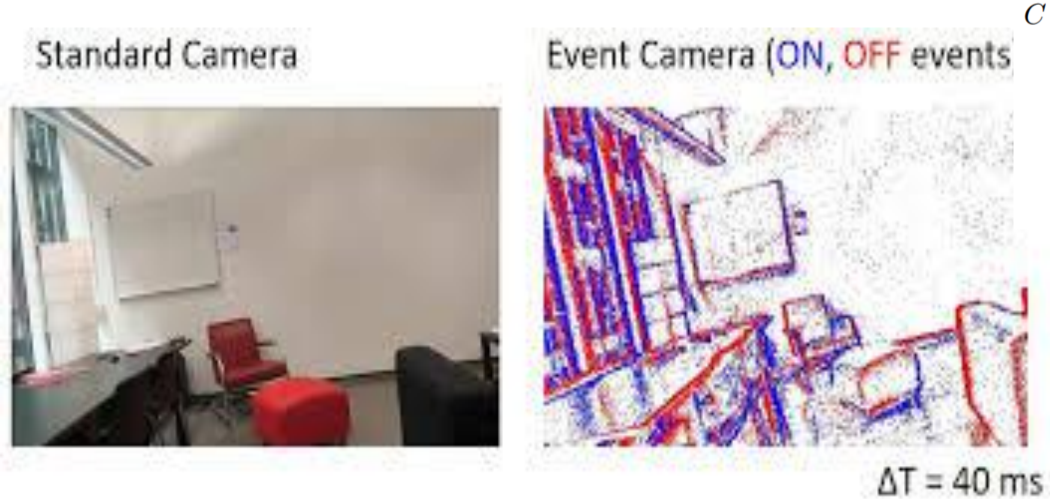
DEFINITION

- What are the event cameras?

DEFINITION (II)

- Each event is made up of four variables:
  - X coordinate
  - Y coordinate
  - Timestamp
  - Polarity

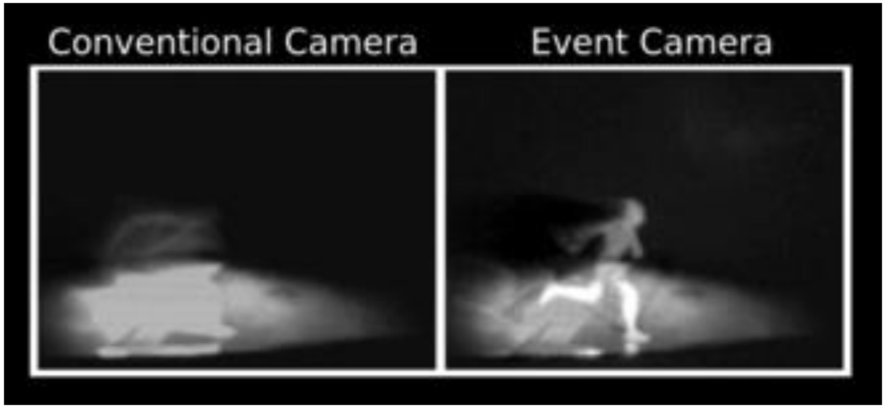




Rebecq, Henri, Daniel Gehrig and D. Scaramuzza. "ESIM: an Open Event Camera Simulator," 2018.


Event-Based Camera vs Standard Camera
Hanme Kim
Robot Vision Group
Imperial College London

- Event cameras:
  - Latency in order of microseconds.
  - Higher dynamic range.
  - No motion blur.
  - No problems when the difference between maximum and minimum brightness levels is very high.



- Original sequence:
  - In some frames, information is lost due to low dynamic range of standard cameras.

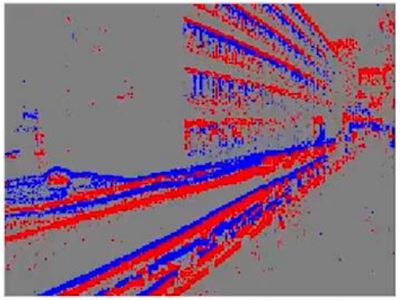
Rebecq, Henri, Ranftl, Rene, Koltun, Vladlen, and Scaramuzza, Davide, "High Speed and High Dynamic Range Video with an Event Camera," 2019.

VIDEO RECONSTRUCTION(II)

- Comparison between the original video and the reconstructed one:
  - Avoids pixel saturation when the illumination difference is very high.

- Comparison between the original sequence and the reconstructed:
  - Event cameras are capable of robust depth estimation, using a single camera model to capture the images.

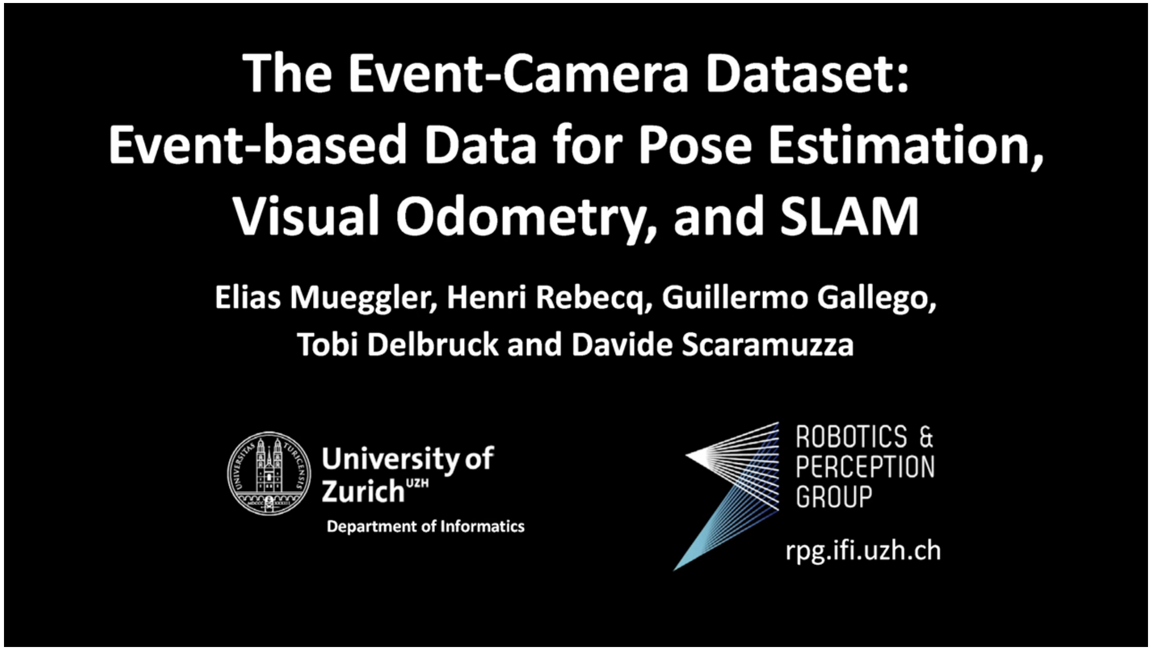
- Simultaneous Localization And Mapping (SLAM) + 3D reconstruction.



Rebecq, Henri, et al. "EMVS: Event-based multi-view stereo—3D reconstruction with an event camera in real-time," 2018.

EVENT CAMERA SIMULATORS

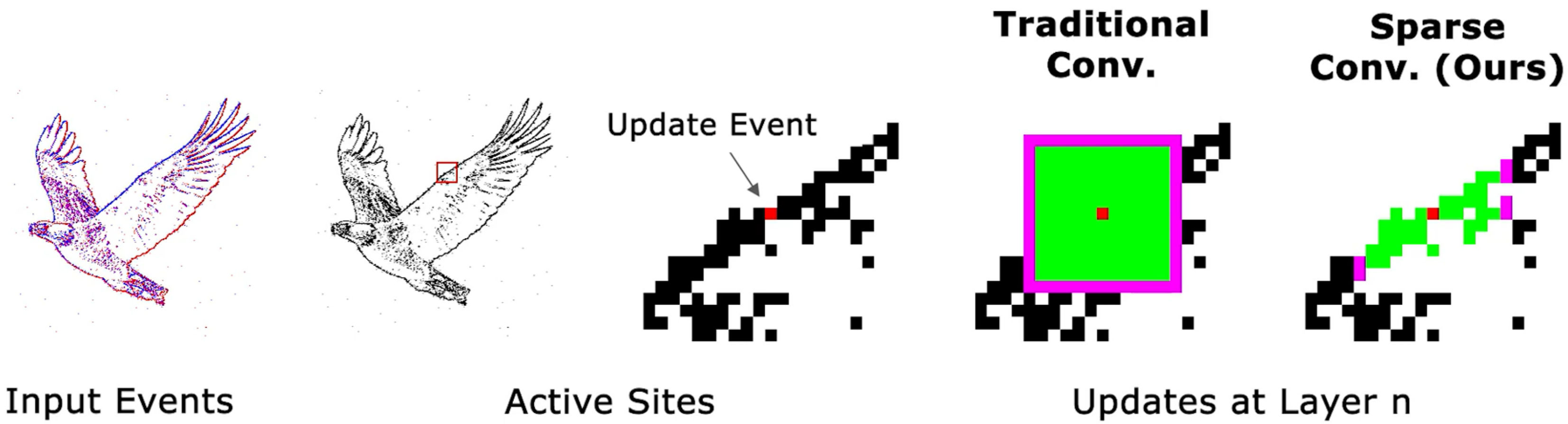
- There is a shortage on the market of events cameras.
- This shortage limits the development of computer vision application algorithms for event signal, specially those that are learning-based.
- To partially solve this situation, simulators are used.



PROBLEM THAT NEED TO BE SOLVED

- Currently, there is no measure the quality of these simulators.
- To solve this situation, the VPULab is working on developing metrics to measure the quality of these simulators.
- To validate these metrics, it is necessary to correlate the distortion measured in event signals with the performance of computer vision applied on distorted event signals.
- We are currently focusing on two applications: **classification** and **object detection**.
  - Using deep learning architectures.

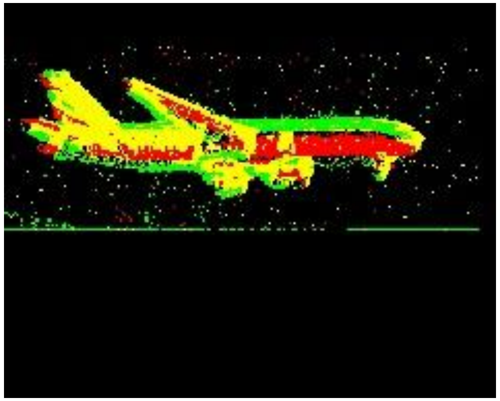
- Their main characteristic is that the used network is based on Sparse Convolution.
  - Sparse Convolution only activates with active pixels, that is, pixels whose value is not zero.
  - It has a lower computational cost than traditional convolution.
  - Due to the fact that events are sparse over the image, this kind of convolution become really useful.

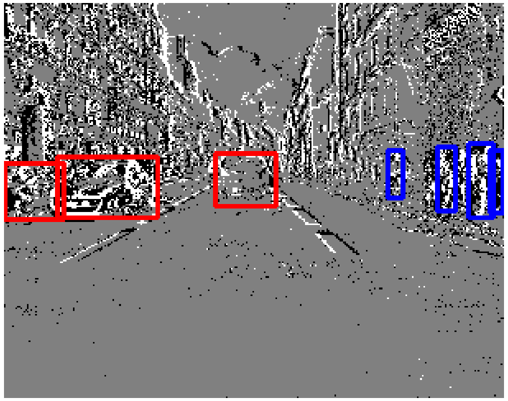


Messikommer, Nico, Gehrig, Daniel, Loquercio, Antonio, and Scaramuzza, Davide, "Event-based Asynchronous Sparse Convolutional Networks," 2020.

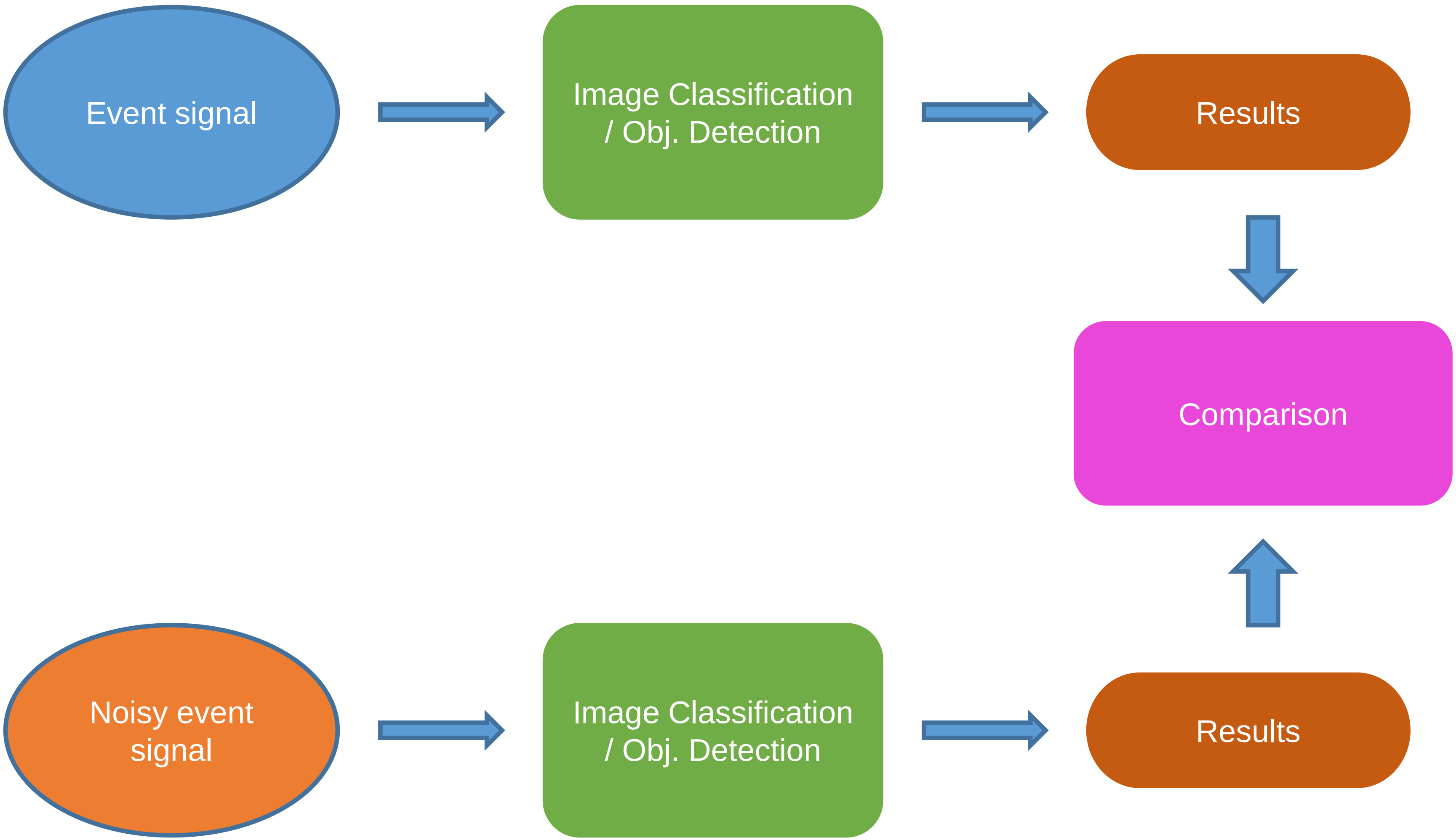
DATASET

- Image Classification:
  - Classification NCaltech101 (101 classes).
  - NCars (Car/No car).
- Object detection:
  - Object Detection NCaltech101 (101 classes).
  - Prophesee Gen1 Automotive (Car/Pedestrian).

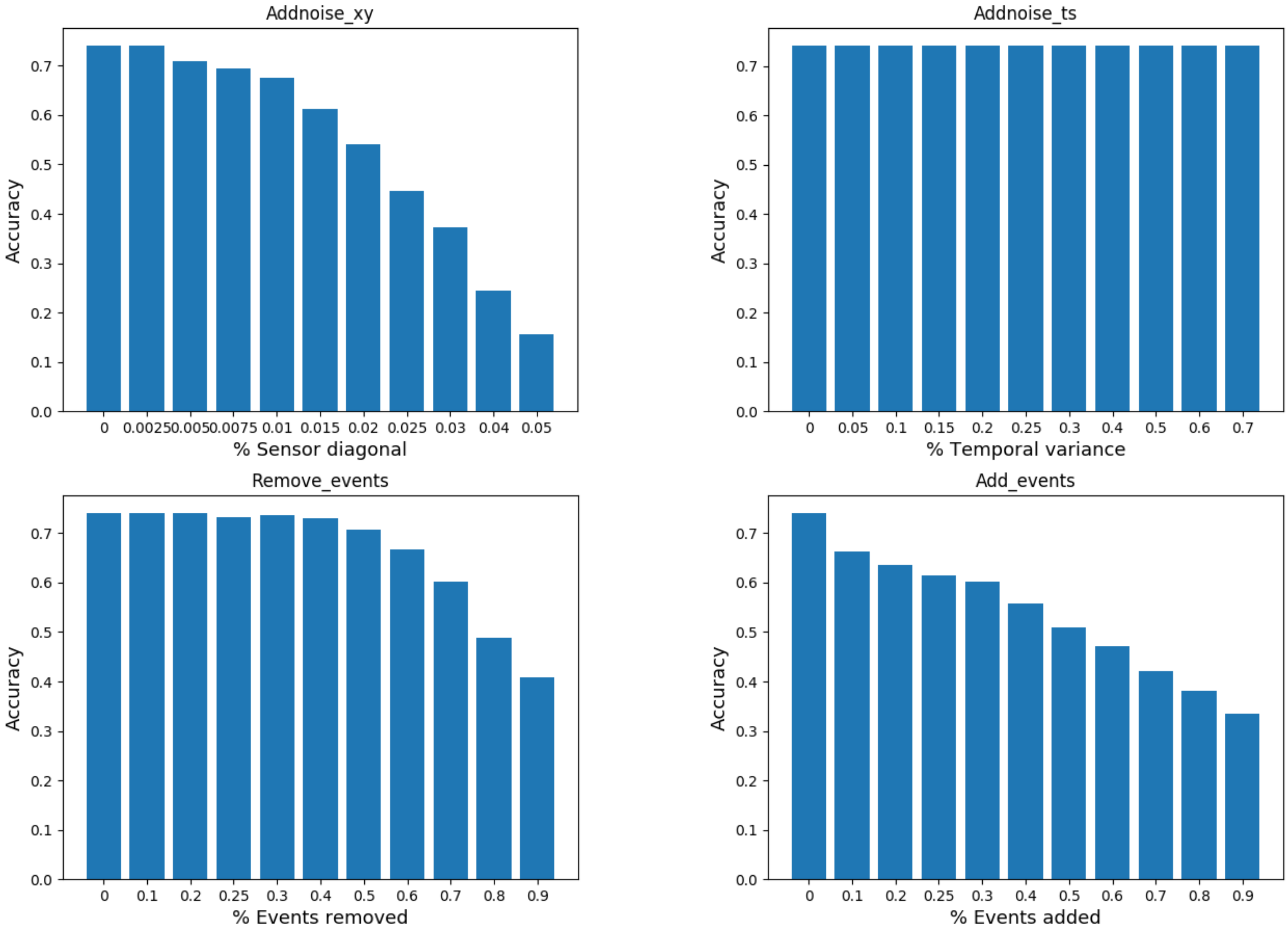




METHODOLOGY



- We are going to show the results that we have obtained in the task of image classification when we use the NCaltech101 dataset.
- A total of four differents types of noises have been evaluated:
  - Spatial noise: Modifies the spatial values of the events.
  - Temporal noise: Modifies the temporal values of the events.
  - Remove events: Removes events from the event sequences.
  - Add events: Creates random events and adds them to the event sequences.



Addnoise_xy
Accuracy
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0 0.00250.0050.0075 0.01 0.015 0.02 0.025 0.03 0.04 0.05
% Sensor diagonal
Addnoise_ts
Accuracy
0 0.05 0.1 0.15 0.2 0.25 0.3 0.4 0.5 0.6 0.7
% Temporal variance
Remove_events
Accuracy
0 0.1 0.2 0.25 0.3 0.4 0.5 0.6 0.7 0.8 0.9
% Events removed
Add_events
Accuracy
0 0.1 0.2 0.25 0.3 0.4 0.5 0.6 0.7 0.8 0.9
% Events added

CONCLUSIONS

- Compared to traditional cameras, event cameras have lower latency (microseconds) and higher dynamic range.
- These two factors prevent us from suffering from motion blur when we are capturing moving images, and also, due to the higher dynamic range, pixel saturation is avoided when the difference in brightness in a scene is very high.
- From the results we have obtained:
  - Spatial noise is the most influential.
  - Temporal noise does not affect the accuracy of the algorithm.
  - Remove events has influence when 40%/50% of events are removed.
  - In add events, the higher the percentage of events we add, the worse the results we get.

- We are currently working on the analysis of the quality of event simulators based on the effect of noise in computer vision applications.

Any

Question