Real world application of event-based end to end autonomous driving

Yuxuan Chen
Mentors: Dr. Igor Gilitschenski, Alexander Amini
Motivation

Autonomous driving cars need to handle a wide range of scenarios:

- Night-time Driving
- No Lane Markings
- Rainy Weather

[Learning steering bounds for parallel autonomous systems, Amini et al.]
How do they do it?
Autonomous Driving Pipeline

Separate problem into smaller sub-modules, tackle each independently

Sensor Fusion
• What’s happening around me?

Detection
• Where are obstacles?

Localization
• Where am I relative to the obstacles?

Planning
• Where do I go?

[Learning steering bounds for parallel autonomous systems, Amini et al.]
End-to-end Learning

Learn the control directly from raw sensor data

Sensor Fusion
• What’s happening around me?

Learned Model
Underlying representation of how humans drive

Actuation
• What control signals to take?

[Learning steering bounds for parallel autonomous systems, Amini et al.]
PilotNet

Learn the steering directly from pixel values

[End to end learning for self-driving cars, Bojarskiet al.]
Problem with RGB cameras

Dynamic Range

Motion blur

Latency
What are event-based cameras

Novel bio-inspired sensors that capture motion in the scene
What are event-based cameras

Novel bio-inspired sensors that capture motion in the scene

[Event-based Cameras: Challenges and Opportunities, Scaramuzza et al.]
What are event-based cameras

Novel bio-inspired sensors that capture motion in the scene

Benefits:
• Low latency (~ 1 microsecond)
• No motion blur
• High dynamic range (140 dB instead of 60dB)
What are event-based cameras

Novel bio-inspired sensors that capture motion in the scene

Benefits:
- Low latency (~1 microsecond)
- No motion blur
- High dynamic range (140 dB instead of 60 dB)

Challenges:
- Data format of events
- Monochromatic
- Low resolution

DAVIS240 from Inivation.com
Our Goal

Use an event camera to drive a car in real time
Related Work: Frame-based models

Event frame
R: positive
G: negative

Network Architecture
Resnet (CNN) → FC → Steering Angle
Related Work: Frame-based models

Problems:
- passive training not tested on a real vehicle
- unable to capture the whole scene at low speed

[Event-based Vision meets Deep Learning on Steering Prediction for Self-driving Cars, Manqueda et al.]
Our proposed model

Thoughts: Augment the event-based model with inputs from a traditional RGB camera, so that the combined model perform at least as well as the best of the RGB-based and event-based models.
Our proposed model

Thoughts: Augment the event-based model with inputs from a traditional RGB camera, so that the combined model perform at least as well as the best of the RGB-based and event-based models.
Comparison between the three models

**PilotNet**

RGB camera 66 x 192 x 3

**Maqueda et al.**

Asynchronous Events $e = (x, y, t, p)$

2-channel integration over 33 ms

180 x 240 x 2

**Ours**

RGB camera 66 x 192 x 3

Asynchronous Events $e = (x, y, t, p)$

1-channel integration over 33 ms

90 x 240 x 1
Experiment Vehicle Setup
Experiment Dataset

2 hours of human driving around Boston on urban roads
Supervise on curvature (1 / radius)
Experiment Metrics

Given ground truth value $\alpha$ and prediction value $\hat{\alpha}$

\[
RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (\hat{\alpha}_j - a_j)^2}
\]

Rooted Mean Square Error

\[
EVA = 1 - \frac{\text{Var}(\hat{\alpha} - \alpha)}{\text{Var}(\alpha)}
\]

Explained Variance
The original event-based model performs the best, but ROI-cropping and 1-channel integration decreased its performance.

Our model, which could be seen as a mixture PilotNet and Maqueda et al. with both ROI cropping and 1-channel integration, indeed perform better than either of them.
Experiment result on real cars

<table>
<thead>
<tr>
<th>Model</th>
<th>autonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>PilotNet [15]</td>
<td>66%</td>
</tr>
<tr>
<td>Maqueda et al. [18]</td>
<td>0%</td>
</tr>
<tr>
<td>Ours</td>
<td>45%</td>
</tr>
</tbody>
</table>

Metric:

\[
\text{autonomy} = (1 - \frac{\text{(number of interventions)} \cdot 6 \text{ seconds}}{\text{elapsed time} \ [\text{seconds}]}) \cdot 100
\]
Discussion

PilotNet also uses Imitation learning, so why does it work better?
Discussion:

Challenges

- Event-based cameras provide structure of the scene and the motion of the camera
- The model turns out to predict the existing motion of the car rather than learning how to drive
Discussion:

Challenges
- Event-based cameras provide structure of the scene and the motion of the camera
- The model turns out to predict the existing motion of the car rather than learning how to drive

Potential solutions for the future
- Use Deep Reinforcement Learning for the model to learn the correct causation
- Work on a event-based simulation platform
Thank you! Questions?

- My mentors: Dr. Igor Gilitschenski and Alexander Amini
- Prof Daniela Rus, Distributed Robotics Lab, MIT CSAIL
- MIT PRIMES
- My parents