16-833: Robot Localization and Mapping (Spring 2019)
Project Proposal
Fusing RGB Keyframes and Event-Based Features for Deep Visual Odometry

Members: Ganesh Iyer (giyer), Abhay Gupta (abhayg), Suhit Kodgule (skodgule)

March 18, 2019

1 Impact

Although various approaches currently exist that combine RGB and event-based cameras for various tasks, (such as Optical Flow, Steering Prediction and Camera Localization), none of the methods reliably leverage the versatility of deep learning to achieve these results, especially with the recent wave of sensor-fusion research being done with the help of this paradigm. In this regard, we showcase a deep visual odometry framework that integrates trajectories to maintain local consistency between trajectory outputs from faster event-based streams and slower RGB keyframes. We also hope to exhibit bundle adjustment at loop closures and produce a semi-dense map output.

2 Novelty

The following contributions will be made in this research:

- A novel representation that allows for combining the use of event-based frames in deep learning along with RGB frames.
- A deep neural network that reliably fuses high-speed event-based pose transforms with poses obtained from RGB keyframes.
- An ablation study (or explainability map) to understand what aspects of the data and frames does the network focus on to predict odometry outputs.

3 Methods from Class

The topics from class being used are

- Rigid-Body Transformations and Manifolds; specifically, differentiable conversion from $SE(3)$ to $se(3)$.
- Since this is essentially a Visual Odometry framework, we hope to demonstrate mapping using Bundle Adjustment (BA) at loop closure.

4 Success Metrics

There are three key metrics which we will be evaluating our results against

- Absolute Translation Error (ATE)
- Relative Pose Error (RPE)
- L2 distance of $SE(3)$ Error

Our algorithm will essentially claim better local performance for trajectory estimates.

5 Key Technical Challenges

The two main technical challenges which we will be facing are

- We need to know how to fuse the poses from RGB keyframes with the poses from the event-based stream (sensor-fusion).
- We will need to decide on a good representation for event-based frames using deep learning.

One key challenge that we are aware of is that it is hard to maintain temporal consistency over long sequences of frames using deep learning approaches.

6 Software and Datasets

The primary dataset that we will be using for this project is the 'Multi-Vehicle Event Camera Stereo Dataset (MSVEC)' [1]. Another dataset which we will be using for testing our methodology is 'The Event-Camera Dataset and Simulator: Event-based Data for Pose Estimation, Visual Odometry, and SLAM' [2]. We will be using PyTorch (0.4.1) and Python (3.6) for our primary development.

7 Proposed Timeline

The proposed timeline for the project will be

- 03/31 - Preparation of data
- 04/07 - Design of algorithm
- 04/15 - Model reiteration - training/testing
- 04/22 - Benchmark Evaluations and Results
- 05/01 - Report and Presentation

8 References

[1] Zhu, Alex Zihao, et al. "The multivehicle stereo event camera dataset: An event camera dataset for 3D perception." IEEE Robotics and Automation Letters 3.3 (2018): 2032-2039. Link

[2] Mueggler, Elias, et al. "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM." The International Journal of Robotics Research 36.2 (2017): 142-149. Link