
Hi! I Know Where You are Going

Tong Bai
tongb@andrew.cmu.edu

Sai Nihar Tadichetty
ntadiche@andrew.cmu.edu

Abstract

Multiple Object Tracking is a classic but challenging task in the field of computer vision. For general tracking pipeline, people do object detection and align the detected bounding box across frames. During the alignment, beside using appearance feature, most state-of-the-art methods would try to predict the trajectory with motion to boost the accuracy. These predictions could bring further benefit if our application need to take reaction based on future movement, eg. autonomous driving etc. Accurate prediction will earn more time for the motion planning to adapt optimal strategy. In this paper, we proposed an framework using deep learning methods combine with motion prediction to better serve the applications need motion planning based on future movement.

1 Introduction

If we say detection and segmentation is about sensing the world, tracking is telling a story, and further more we want to take precautions. Prediction accompanied tracking long before people adapt machine learning, in traditional vision, researchers using optical flow, and Kalman Filter to do encode the motion information and predict based on the previous motion pattern. Later on, Deep Learning make a revolution in vision and gain state-of-the-art result nearly in every recognition related field. A lot of works trying to borrow the wisdom of deep neural network to solve this classic task. Kim et.al [1] proposed a method under the Multi-hypothesis Tracking(MHT) framework, he tries to use Deep CNN to extract appearance feature, plus Kalman Filter to encode the motion information, and use a tracking tree to get the optimal solution. Further more, Alahi propose a "pure machine learning" approach called Social LSTM [2] which using a LSTM to encode the position information and predict the future trajectory. But both problem are not really suitable for real-time prediction with moving camera. For MHT, the biggest advantage is it allow tracing back to find the optimal solution and correct its most recent mistakes, which also means without knowing the following frames, the current solution might not be optimal. For Social LSTM, the experiments setting are all with static cameras, we don't know how it will work with moving cameras, and the possibility of applying this technique to applications like drone or autonomous vehicle. In this paper, we aiming to experiment and modified the existing tracking/motion prediction methods on real world drone navigating or autonomous driving problems, in order to help accurate motion planning in those applications.

References

- [1] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, "Multiple hypothesis tracking revisited," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4696–4704.
- [2] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.