Final project for CSE576 *Computer Vision*, May 2018

A Real-time UAV-Based Intelligent Tracking System

Hao Xiao, Sujie Zhu, Haotian Zhang
Department of Electrical Engineering (EE)
University of Washington, Seattle, WA

Introduction

Background:
- State-of-art object detectors can achieve real-time detection and keep a high detection accuracy.
- Multi object tracking has been challenge mainly due to noisy detection sets and frequently switches caused by occlusion and similar appearance.
- A real-time tracking system can take us one step further to realizing a fully automatic and highly intelligent security system.

Objective:
- Integrates state-of-art object detection algorithm and multi object tracking algorithm into one real-time tracking system.
- Test the system on Drone streaming video

Data Description

Datasets:
- **PASCAL VOC:** a very popular dataset for building and evaluating algorithms for image classification, object detection, and segmentation. It includes 20 different object, such as people, animal, vehicle and other indoor object.
- **MOT challenge dataset:** a benchmark contains video sequences in unconstrained environments filmed with both static and moving cameras.

Data Processing:
- Configure the Drone and computer for streaming the Drone camera to computer for real-time tracking

Framework

The system can be roughly divided into two mainly part, which are object detection part and multi object tracking part. First, we will capture the image from camera and apply object detection frame by frame. For every object in one frame, we extract the appearance feature and motion feature and send these information into multi object tracking part. Then, we can get the ID of each object.

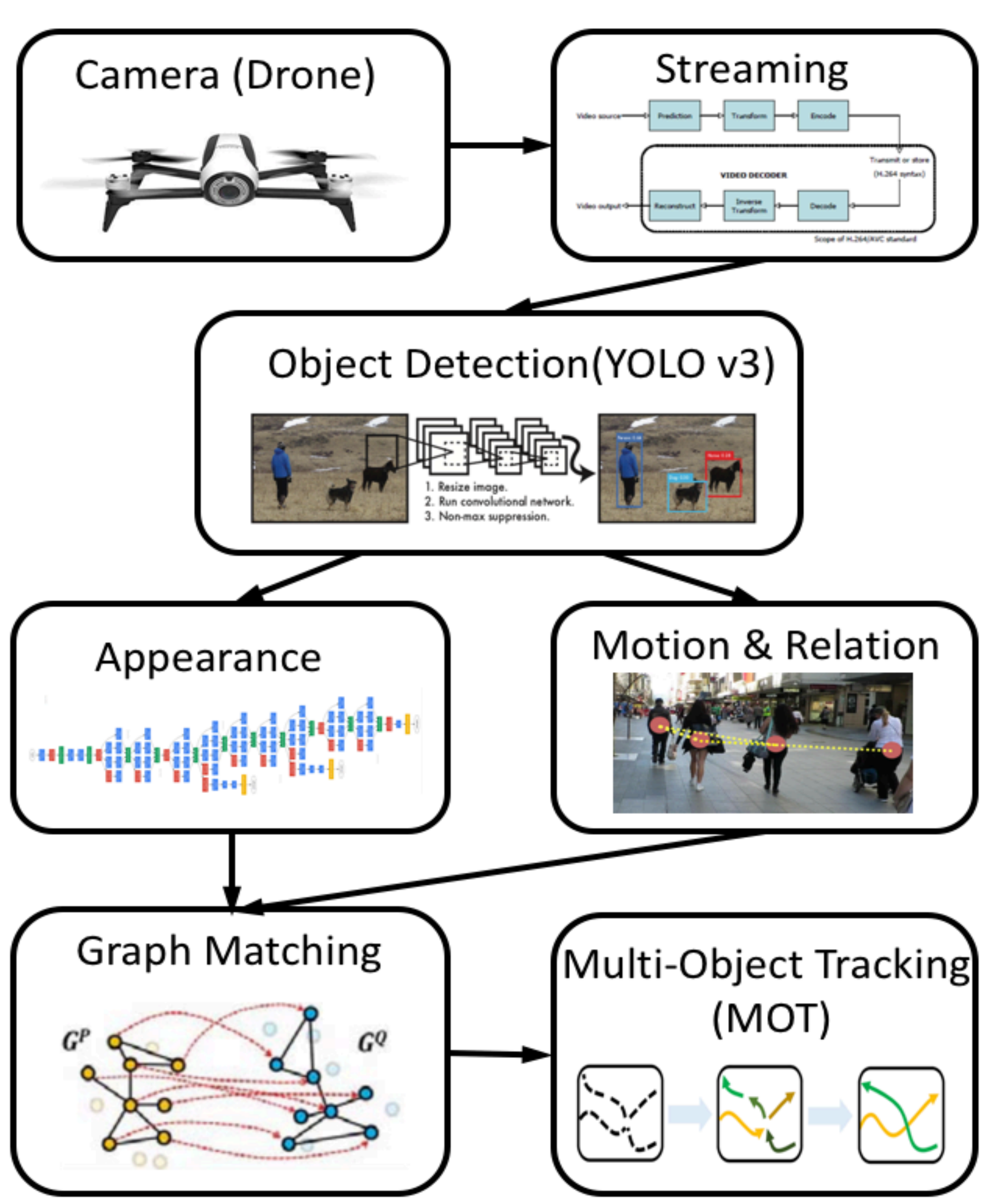

Figure 1. Framework of real-time tracking system

Object Detection & Multi Object Tracking

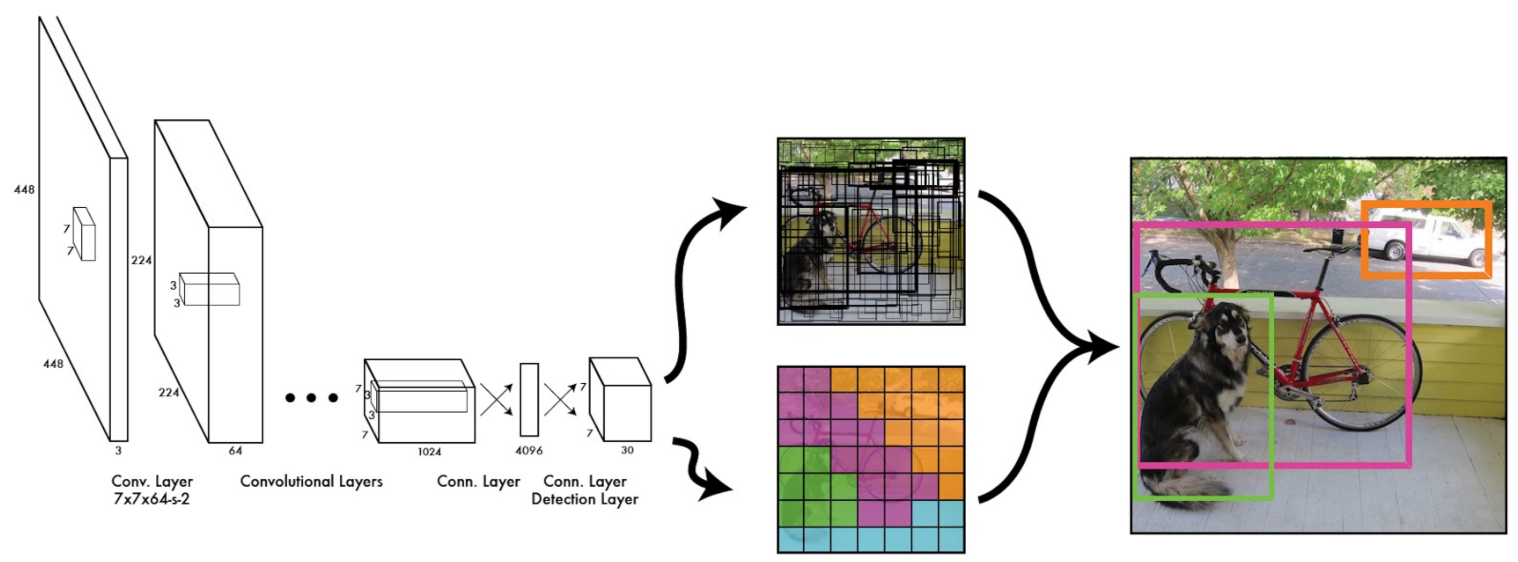

Figure 2. YOLO Detection pipeline

Object detection:

Pre-trained YOLO-v3 on **PASCAL VOC**

Real-time Tracking:

Observation model (object representation)

Appearance:
- traditional SIFT descriptor on each object
- pre-trained GoogleNet on ImageNet
- last FC layer output:1024 dimension vector

Motion:
- Constant velocity motion models/linear motion models:

$$\begin{cases}(x,y)_t = (x,y)_{t-1} + (u,v)_{t-1}\cdot\Delta t + \epsilon_{x,y}\\(u,v)_t = (u,v)_{t-1} + \epsilon_{u,v}\end{cases} \Rightarrow V(d_t^i) = \frac{Ap(T_{t-k}^i) - p(d_t^i)}{k}$$

Interaction:
- Define 5 types of relation: 1) No relation 2) Get closer 3) Get away 4) Occlusion 5) Group

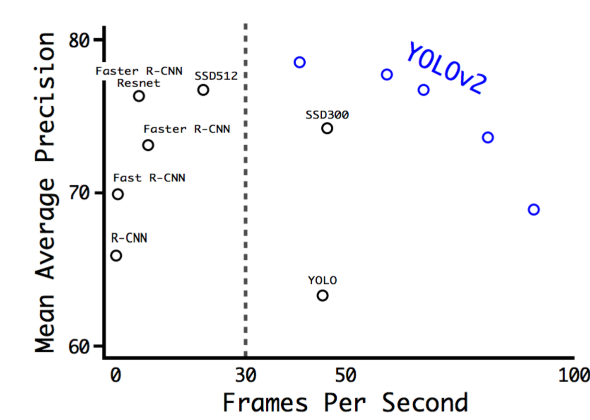

Figure 3. Detection framework performance

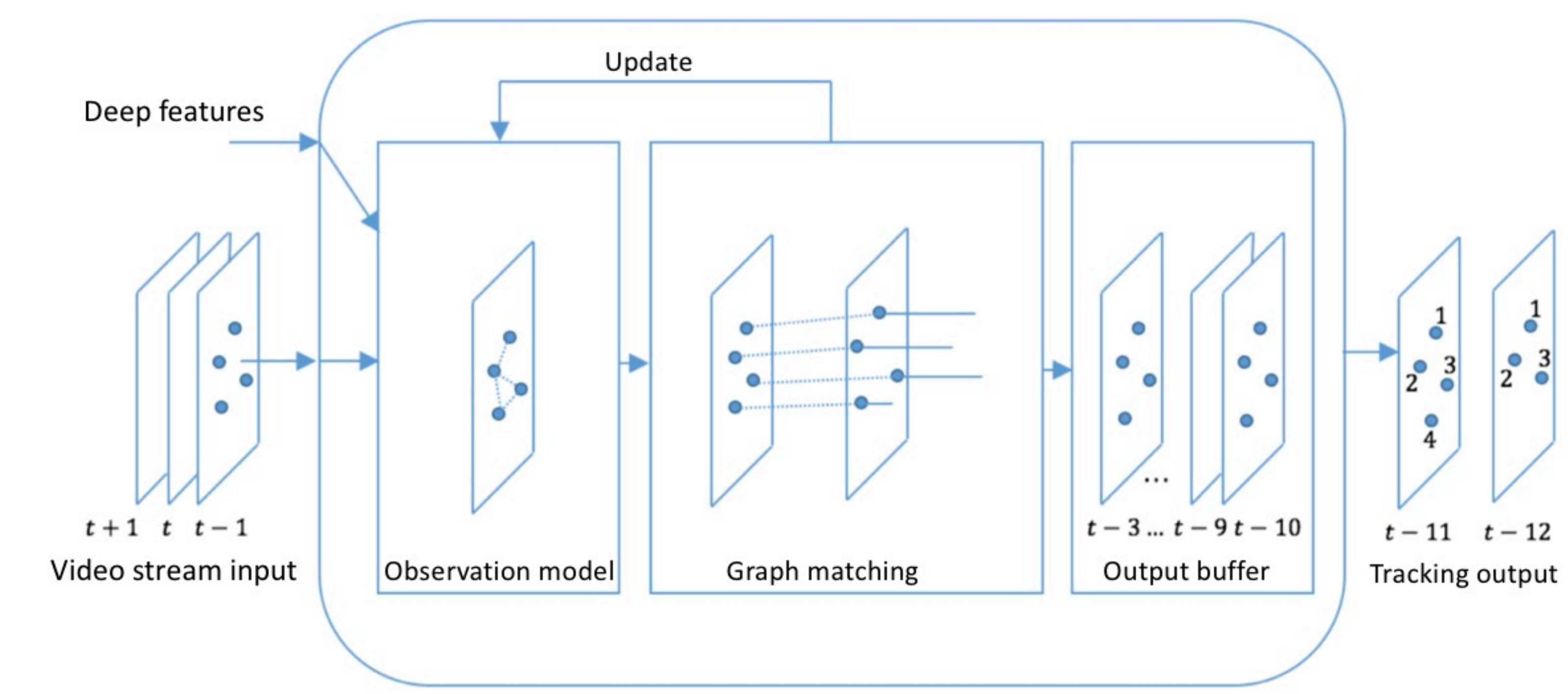

Figure 4. Framework of multi-object tracking(MOT)

Dynamic model (graph matching method)

Graph Definition: $G(V,\ E,\ A)$ and $G'(V',\ E',\ A')$
- Vertex affinity: Appearance & Motion feature
- Edge affinity: Interaction representation

Graph Matching:
- $S(G,\ G',\ y)$ evaluates the similarity between graphs
- Find the best y^* to maximize $S(G,\ G',\ y)$

$$y^* = \underset{y}{argmax}\, S(G, G', y),$$

$$s.t. \begin{cases} y \in \{0,1\}^{nn'}, \\ \sum_{i=1}^{n} y_{ia} \le 1, \sum_{a=1}^{n'} y_{ia} \le 1, \end{cases}$$

Matching Algorithms:
- Traditional Hungarian Algorithm on unweighted bipartite graph for maximum matching

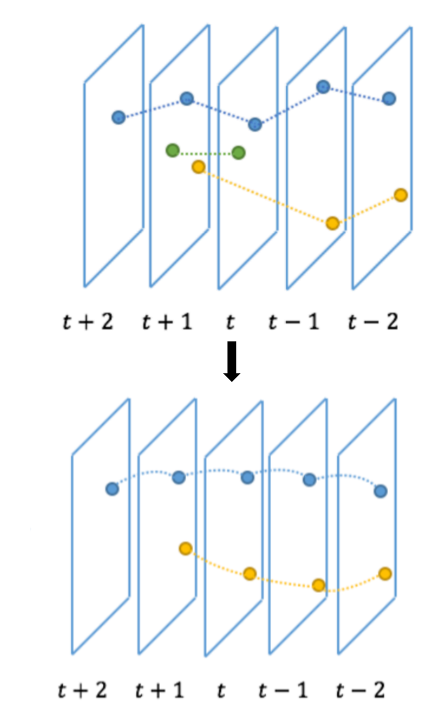

Figure 5. MOT post-processing

Results Visualization and analysis

Figure 6. Visualization of performances on different datasets

MOT Challenge				
Tracker	MOTA (%)	MOTP (%)	ID Matching	Processing Speed (Hz)
NJBst	43.70	73.70	710	39.10
JointMC	35.60	71.90	457	0.60
mLK	35.10	71.50	700	1.00
Our Method	36.60	71.40	700	34.50

Figure 7. Results for some MOT trackers performances vs. Ours

Result Analysis

Figure. 6 shows some clips of our tracker performances on different datasets, including official ETH-Crossing, ADL-Rundle-3 and one video shoot by ourselves.

As shown in the Figure. 7 below, compared to the trackers for MOT Challenge, our tracking algorithm performs quite good not only on the tracking performances but also on the processing speed.

Conclusions & Future Work

- An online multiple object tracking algorithm was implemented;
- The application can be used for UAV surveillance and pedestrian tracking;
- Still, there are some problems, including the frequent ID switch, trajectory fragmentations…
- The performance can be improved by implementing idea of tracking-by-detection, and use camera-self calibration to do 3D tracking (depth information);

References

[1] Wang B, Wang G, Chan K L, et al. Tracklet association by online target-specific metric learning and coherent dynamics estimation[J]. IEEE transactions on pattern analysis and machine intelligence, 2016.
[2] Luo W, Xing J, Zhang X, et al. Multiple object tracking: A literature review[J]. arXiv preprint arXiv:1409.7618, 2014.
[3] Xiang Y, Alahi A, Savarese S. Learning to track: Online multi-object tracking by decision making[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 4705- 4713
[4] Yang B, Nevatia R. Multi-target tracking by online learning of non-linear motion patterns and robust appearance models[C]//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 1918-1925.

**We would like to thank Prof. Shapiro , TA Bindita for great lectures and help.*