Orientation- and Scale-Invariant Multi-Vehicle Detection and Tracking from Unmanned Aerial Videos

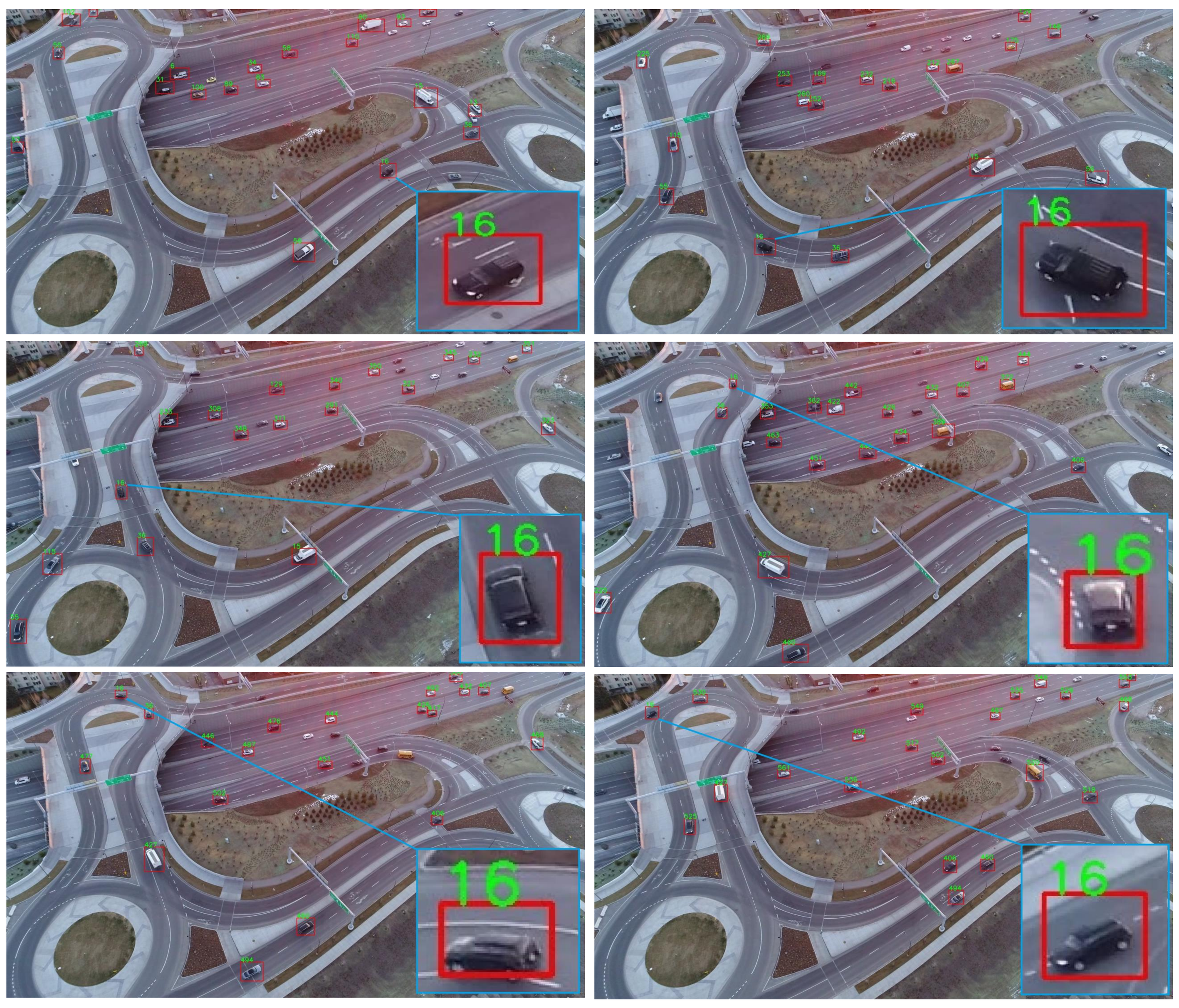
Jie Wang*, Sandra Simeonova, Mozhdeh Shahbazi {wang4@ucalgary.ca}
Department of Geomatics Engineering, University of Calgary, Calgary, AB, Canada

Abstract

Along with the advancement of light-weight sensing and processing technologies, unmanned aerial vehicles (UAVs) have recently become popular platforms for intelligent traffic monitoring and control. UAV-mounted cameras can capture the traffic flow videos from various perspectives providing a comprehensive insight to the road conditions. To analyze the traffic flow from remotely captured videos, a reliable and accurate vehicle detection-and-tracking approach required. In this work, we propose a framework for vehicle detection-and-tracking from UAV videos for monitoring large-scale dense traffic flow in complex road structures. The proposed methodology is tested on a variety of real videos collected by UAVs under various conditions, e.g., in late afternoons with long vehicle shadows, in dawn with vehicles lights being on, over roundabouts and interchange roads where vehicles directions change considerably, and from various viewpoints where vehicles images undergo substantial perspective distortions.

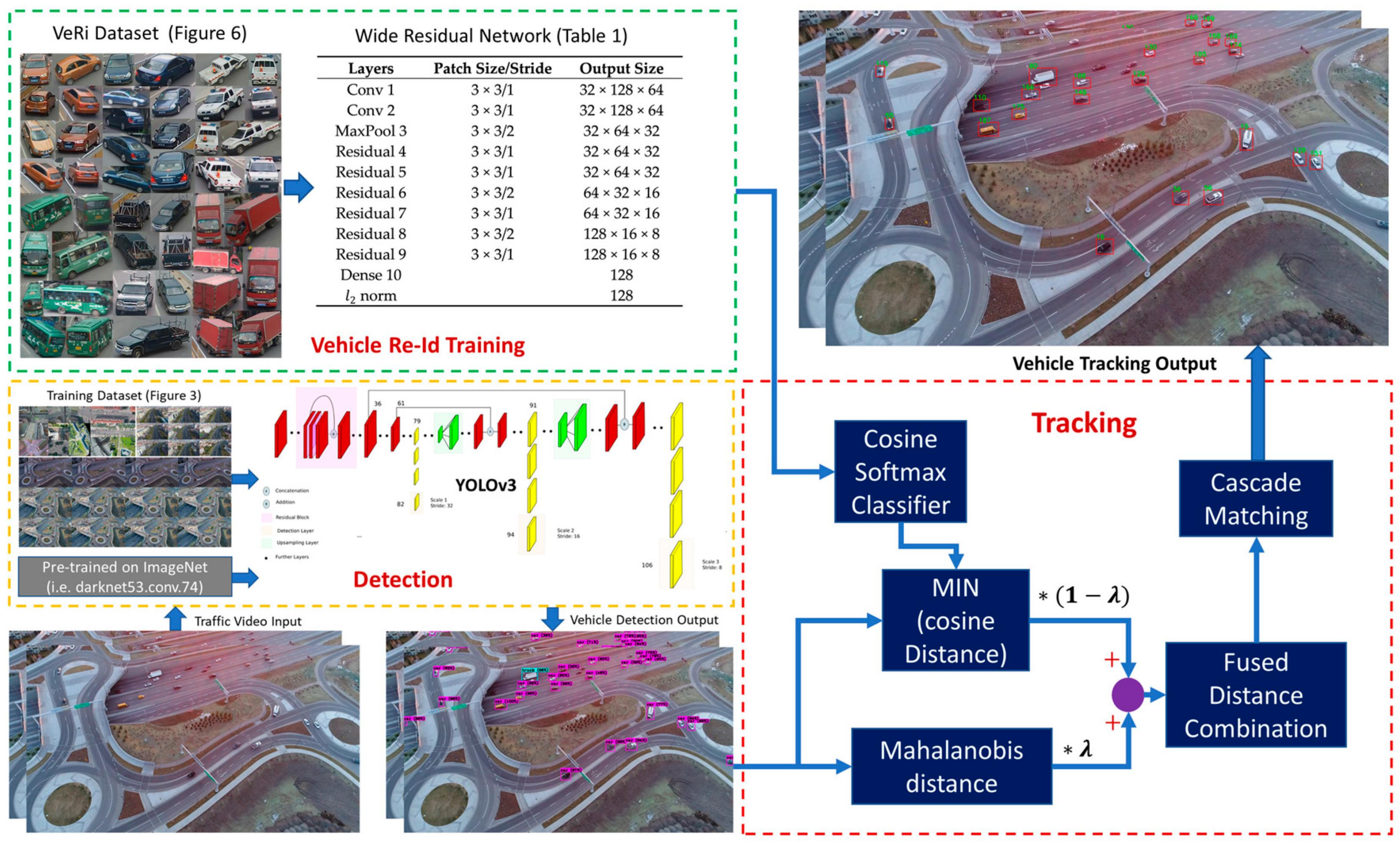
Introduction

- In complex traffic scenarios, such as vehicles moves through multiple roundabout or interchange roads, vehicles maneuvers cannot be tracked completely using traffic videos collected by ground surveillance cameras at fixed positions and orientations.
- UAVs can be complementary platforms to collect traffic videos from various altitudes and viewpoints.
- Challenges: appearance similarities among vehicles are higher due to limited vehicle models and colors, vehicles appearance features are more limited and less distinctive due to smaller pixel size, appearance of the same vehicle varies aggressively due to its changing relative orientation to the camera, frequent identity switching of the tracked vehicles



Methodology

Overview



Detection

- The vehicle detector was fine-tuned on YOLOv3 by a custom labeled dataset together with two other UAV datasets.
- 157 custom labeled images from two videos captured by a DJI Phantom 4 Pro with 2720*1530 resolution at 30 fps.
- The pre-trained weight, darknet53.conv.74 on ImageNet, was used to fine-tune the vehicle detector.

Re-identification Training

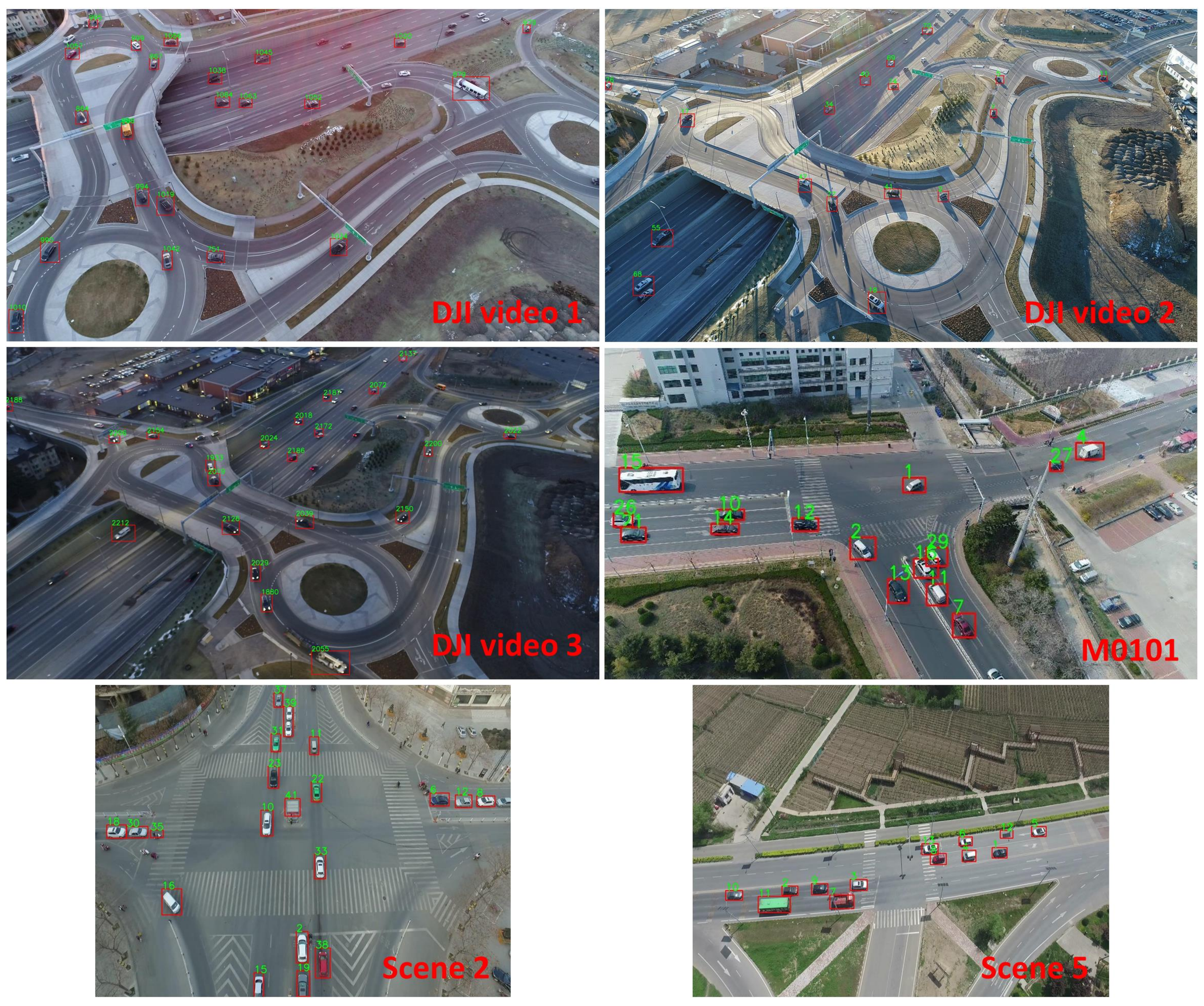
- Deep vehicle re-identification (Re-Id) appearance features were extracted by training a wide residual network with the VeRi dataset.
- The wide residual network was originally designed for multiple people tracking tasks. The VeRi dataset was originally created for vehicle Re-Id purposes across multiple ground surveillance cameras.

Tracking

- Vehicle tracking is conducted following a tracking-by-detection paradigm. A Kalman filtering based motion and Re-Id deep appearance features are integrated for object association in tracking.
- The Squared Mahalanobis distances between the predicted bounding boxes and the new detected ones are calculated as the motion metric as $d_1(i,j) = \left(d_j - y_i\right)^T S_i^{-1}\left(d_j - y_i\right)$
- The minimum cosine distance of the deep appearance features between i and j object as $d_2(i,j) = min\left(1 - r_j^T r_k^{(i)}\right)$
- Two metrics were summed in a weighted form as $c_{i,j} = \lambda d_1(i,j) + (1-\lambda) d_2(i,j)$

Results

- Metrics evaluation: True Positive (TP), False Positives (FP), True Negative (TN), False Negatives (FN), Mostly Tracked (MT), Mostly Lost (ML), Identity Switching (IDSW), and Multiple Object Tracking accuracy (MOTA).
- The vehicle tracking method performed accurately with average MOTA being 81.5%. The frequent identity switching problem of the multiple-object tracking tasks was solved well, resulting in an average of 1 identity switches over 500 tracked vehicles.



Metrics evaluation results

Video	Total #	TP↑	FP↓	TN↑	FN↓	MT↑	ML↓	IDSW↓	MOTA↑
DJI video 1	14496	11949	0	0	2456	52.63%	0%	24	82.89
DJI video 2	5413	3529	299	0	1298	46.15%	0%	90/41	68.83/69.74
DJI video 3	15120	7319	38	0	2543	44%	4%	45/43	82.63/82.65
M0101	558	4441	8	0	610	76.19%	0%	8/5	88.74/88.79
Scene 2	10027	9385	783	0	597	90.90%	0%	13/6	86.11/86.18
Scene 5	16061	11305	161	0	3227	68.57%	5.7%	22/19	78.77/78.78

The testing video results available at:
https://github.com/jwangjie/UAV-vehicle-tracking

@ Deep Learning and Reinforcement Learning Summer School 2019