AI City Challenge 2019 – City-Scale Video Analytics for Smart Transportation

Ming-Ching Chang [1] Jiayi Wei [2] Zheng-An Zhu [3] Yan-Ming Chen [3] Chan-Shuo Hu [3] Ming-Xiu Jiang [3] Chen-Kuo Chiang [3]

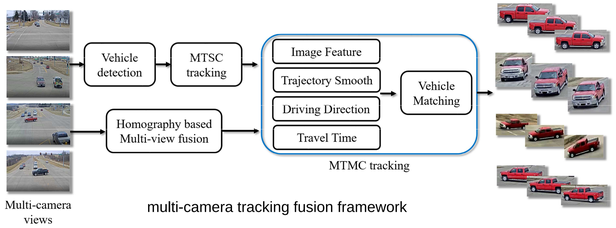
[1] University at Albany – SUNY, NY, USA [2] Beijing University of Posts and Telecommunications, China [3] National Chung Cheng University, Taiwan

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Track 1: City-Scale Multi-Camera Vehicle Tracking

Introduction

Track 1 challenge is on city-scale vehicle tracking from 40 cameras spanning a mid-size US city (the CityFlow dataset). We propose a multi-camera fusion framework combining: (1) single-cam tracking-by-detection, (2) tracklet association, (3) vehicle Re-ID, and (4) geographical/site calibration knowledge into a single optimization framework.



multi-camera tracking fusion framework

Method

Homography based multi-view fusion

- Map pixels from multi-cam videos into world coordinates.
- Automatic ROI selection for each video to mask out non-groundplane regions.

Multi-Target Multi-Camera Tracking

Spatial-temporal fusion based on:

- Tracklet association (for overlapping views)
- Vehicle Re-ID (for non-overlapping views)
- Fusion algorithm: iterative greedy association of a pair of vehicle tracks (T_i, T_j) in cameras (C_i, C_j) respectively, by minimizing function $L(T_i, T_j)$ characterizing dissimilarity.
- $L(T_i, T_j)$ consists of 4 terms: **image feature loss** L_f, **trajectory continuity loss** L_c, **driving direction loss** L_d, **and travel time loss** L_t in weighted combinations:

$$L(T_i, T_j) = \omega_f L_f + \omega_c L_c + \omega_d L_d + \omega_t L_t$$

Experimental Results

AIC'19 evaluation: $S_1 = F1_{id}$ score 0.1634 (rank 17 out of 22 teams).

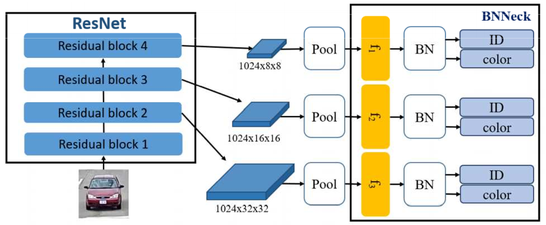
Track 2: City-Scale Multi-camera Vehicle ReID

Introduction

- Goal is to match the identities of detected/tracked vehicles provided in their *algorithmic generated* bounding boxes across cameras.
- Challenge lies in large intra-class variabilities (greater than inter-class variabilities).
- We propose **Pyramid Granularity Attentive Model (PGAM)** to effectively extract both coarse and fine-grained features, inspired from:
  - Region-Aware Deep Model (RAM), Multiple Granularities (MGN).
- Training Tricks / Strategies to retain fine-grained discriminability:
  - Random Erasing Augmentation, Center Loss, BNNeck

Pyramid Granularity Attentive Model

- Multi-branch design in previous ReID works only works for well-cropped vehicles for ReID. Here we use a multi-scale model to better extract both coarse and fine-grained features.

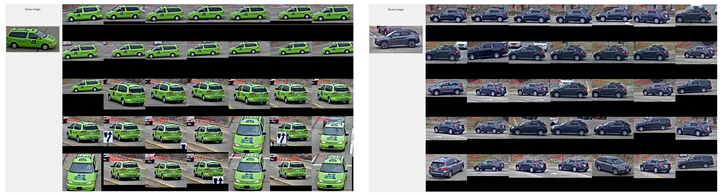


- Given an input vehicle image, we use the last three residual blocks of ResNet to obtain image features, each with 1024 channels.

Experimental Results

Metric: Top-K mAP (K=100) on CityFlow Re-ID dataset.

Name	Status	mAP	CMC_1	CMC_5	CMC_10	CMC_15	CMC_20	CMC_30	CMC_100
pyramid_1	Success	0.2953	0.4449	0.5105	0.5428	0.5970	0.6378	0.6996	0.7785



Track 3: Traffic Anomaly Detection

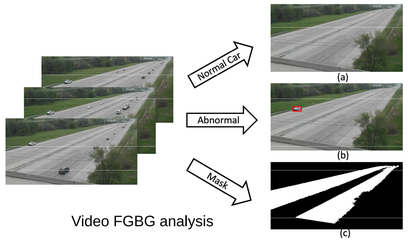
Introduction

- High impact: make traffic systems smarter and safer in reducing response time.
- **Anomaly**: crashes, emergency stops related to *stalled vehicles*.
- Iowa Department of Transportation (DOT) dataset:
  - Large variabilities in weather conditions, snow, day/night, glare, camera vibrations, video transmission artifacts, …
  - 100 training and 100 test videos, each 15 mins, 30 FPS, 800*410.
  - Training set labeling: anomaly (begin, end) time.
  - Traffic light stops, cars away from main roads, freeze video frames are not counted as abnormal.

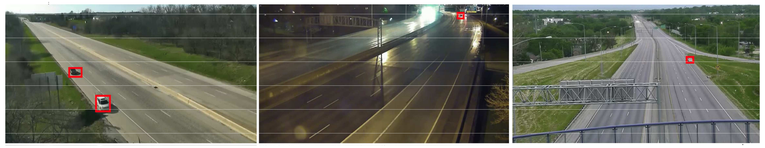
Method

Start with detecting stalled vehicles after a foreground/background (FGBG) filtering, followed by a back-tracking of all detected stalled vehicles to accurately localize event occurrence time.

- Use MOG2 (OpenCV) foreground/background analysis.
- Use optical flow to detect road mask, video scene changes and freeze video frames.
- Detect vehicle in background frames using a DNN model combining design of R-FCN, Feature Pyramid Network (FPN), Focal Loss.
- Rule out unsittable vehicle detections using classifier scores, road mask, and freeze frames.
- Back-tracking to locate the starting time of each abnormal vehicle using DaSiamRPN.



Video FGBG analysis

Experimental Results

With 0.7143 F1 score and 6.1057 RMSE, we achieved 0.6997 S3 score and ranked 6th (out of 23 teams) in the final leaderboard. Example results shown below: