POSTER P21912 | CONTACT NAME Kazuyuki Miyazawa: kazuyuki.miyazawa@dena.com

Geolocation of Traffic Lights and Signs Using Dashcam: Towards Low-Cost Map Maintenance

Kazuyuki Miyazawa, Kosuke Kuzuoka, Wensheng Ran, and Hirohito Okuda (DeNA Co., Ltd., Japan)

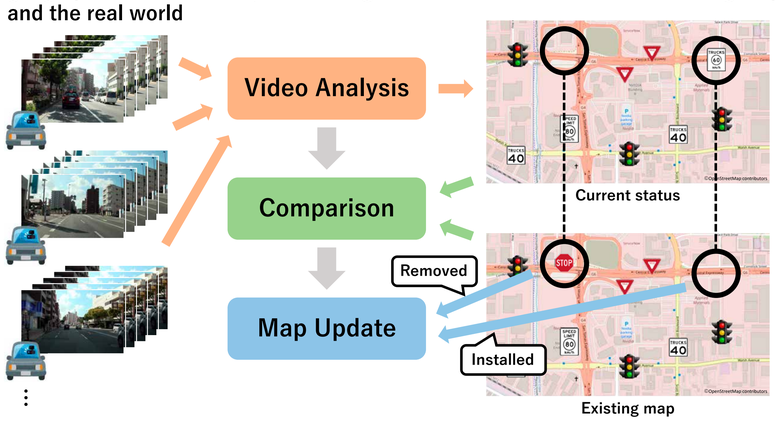
Background

- Maps for autonomous driving era such as HD maps need to be maintained as close to the current real world as possible (i.e., difference between the existing maps and the real world should be resolved immediately)
- Collecting data for updating existing maps manually or by using special mobile mapping systems with numerous sensors is quite costly and time consuming
- Need to develop a (near) real-time system that can point out the differences automatically using data collected by low-cost sensors like consumer cameras

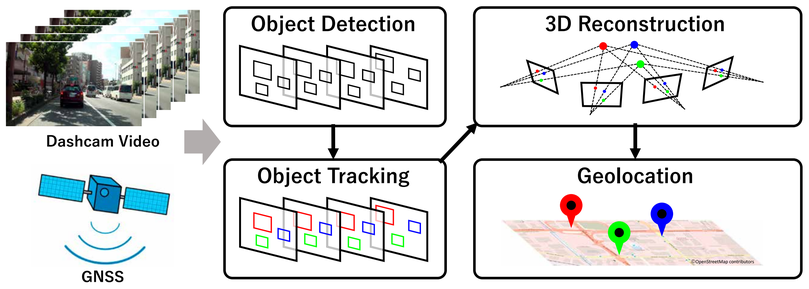
Our Goal

Develop technologies and systems that enable us to easily maintain maps without employing many workers or special vehicles

- Installing a dashcam is becoming quite popular and there are many services that automatically upload the videos to cloud system mainly for data preservation
- Such videos contain rich information for map maintenance, and the current advanced computer vision technology can be utilized to extract the information
- Key technologies are detecting target objects from videos, estimating their 3D geographical coordinates, and finding the differences between the existing maps and the real world



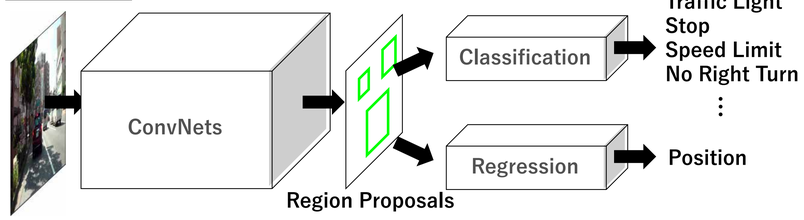
Overview



- Currently developing a basic pipeline that geolocates each traffic light and sign using only videos from low-cost dashcams and GNSS information (i.e., Video Analysis part in "Our Goal")
- The pipeline consists of object detection, object tracking, 3D reconstruction, and geolocation, and most of them can benefit from GPU acceleration
- Collect dashcam videos with GNSS information by running more than 25,000 miles on Japanese roads in various areas and annotate the videos by our in-house annotators and tools
- Experimental evaluation using the dataset demonstrates that our system can detect objects with more than 85% accuracy, and can estimate their locations with less than 10m error under existence of GNSS error

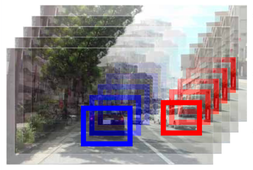
Traffic Lights/Signs Detection and Tracking

Detection



- Our object detector is based on Faster R-CNN [1] with a classification head that can classify more than 100 types of Japanese traffic lights and signs
- Object classes in dashcam videos is highly imbalanced, so data augmentation is essential and the policy for such augmentation is tuned carefully
- Not only apply simple image transformations, but also utilize synthesized images to alleviate the low-data problem

Tracking

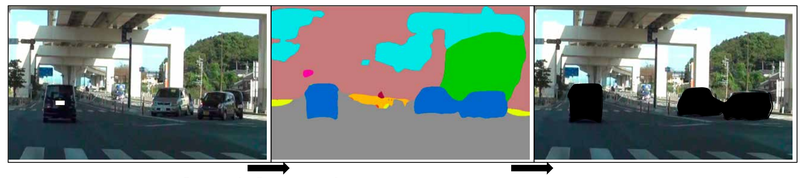


- Since the bounding boxes obtained by our detector are reliable, each object can be tracked by simple algorithm
- Evaluate only IoU of bounding boxes between frames, and continue tracking if a certain IoU threshold is met
- Our simple tracker does not need to access any image information and runs much faster than frame rate

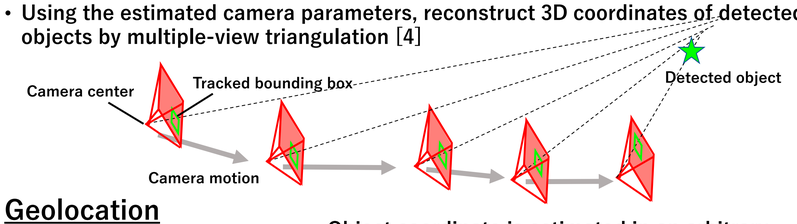
3D Reconstruction and Geolocation

3D Reconstruction

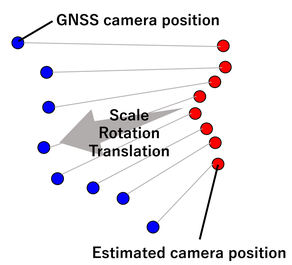
- For 3D reconstruction of the detected objects, employ Structure from Motion (SfM) [2] to estimate the intrinsic parameters of cameras and 3D camera motion
- Since SfM assumes that the scene is static, moving objects degrade the performance of 3D reconstruction
- Apply semantic segmentation [3] to each frame, and mask all the movable objects to improve the performance of SfM



- Using the estimated camera parameters, reconstruct 3D coordinates of detected objects by multiple-view triangulation [4]



Geolocation



- Object coordinate is estimated in an arbitrary coordinate system in SfM, so need to convert it to the geographic coordinate system based on camera positions from GNSS
- Coordinate conversion can be represented by similarity transformation (i.e., scale, 3D rotation, and 3D translation)
- Employ RANSAC (RANdom SAmple Consensus) to robustly estimate the transformation parameters with the existence of outliers

Experiments

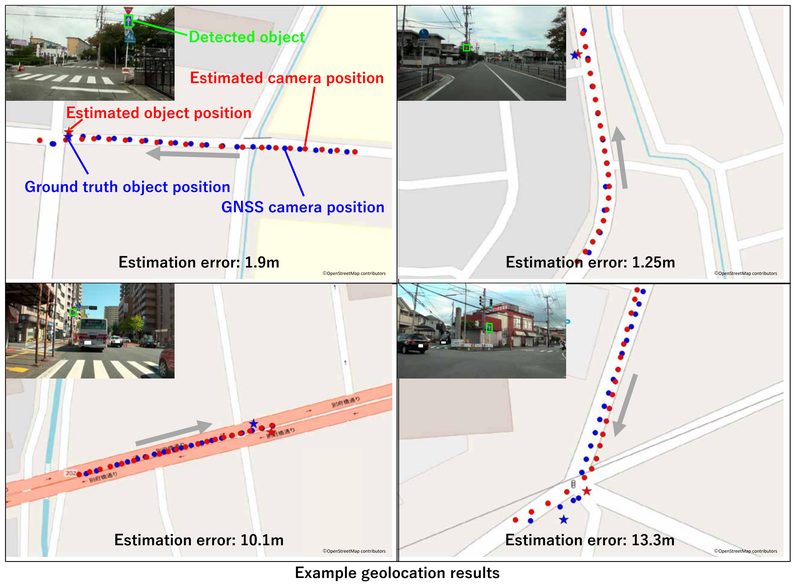
Dataset

- Collect dashcam videos with GNSS information by running more than 25,000 miles on Japanese roads in various areas (city, country, suburb etc.)
- Annotate traffic lights and signs in each frame, and manually geolocate each object to collect their ground truth geographic coordinates

Results

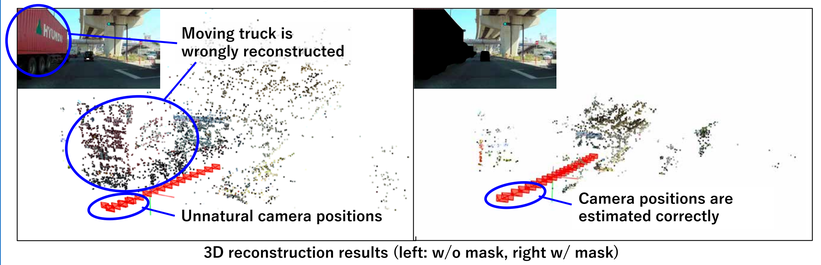
- For object detection, our detector can achieve more than 85% recall
- For geolocation, the average distance between estimated coordinates and ground truth (i.e., estimation error) is about 10m including GNSS error*

* It is said that GNSS single-point positioning error is about 10-20m.



Example geolocation results

- Masking moving objects by semantic segmentation is highly effective to improve the 3D reconstruction accuracy in SfM



3D reconstruction results (left: w/o mask, right w/ mask)

Conclusions

- Develop the basic pipeline for detecting traffic lights and signs and estimating their geographical coordinates using only dashcam videos and GNSS information towards low-cost map maintenance without special systems
- Build our own dataset that includes a large amount of dashcam videos with bounding box annotations and ground truth geographical coordinates
- Achieve more than 85% recall and about 10m coordinate estimation error

References

[1] S. Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," NeurIPS 2015.

[2] J. L. Schonberger et al., "Structure-from-Motion Revisited," CVPR 2016.

[3] K. Sun et al., "High-Resolution Representations for Labeling Pixels and Regions," CoRR, 1904.04514, 2019.

[4] R. I. Hartley et al., "Multiple View Geometry in Computer Vision," Cambridge University Press, 2004.