The Pulse of Allegheny County and Pittsburgh

FINAL RESEARCH REPORT

José M F Moura (PI), Evgeny Toropov,
Joya Deri, Satwik Kottur

Contract No. DTRT12GUTG11

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

1. Problem Statement

Cities are increasingly equipped with low-resolution cameras. They are cheap to buy, install, and maintain, and thus are usually the choice of departments of transportation and their contractors. Pittsburgh or New York City have networks of hundreds of cameras. Video from some of these cameras is publicly accessible in real time.

In this project, we addressed the problem of building a traffic model for parts of the roads visible from publicly accessible cameras. In particular, our end goal is to build a model capable of detecting different types of vehicles in images in various weather conditions and times of the day except night. Models learn different appearance of vehicles as seen from different viewpoints. A major difficulty with any type of analysis like this is the need for large amounts of training data. In our case, it is easy to collect unlabeled data from publicly available low-resolution low-framerate cameras in Pittsburgh or NYC (figure 1).

Fig. 1: typical images from NYC cameras

Some contractors from industry recently made substantial investments into the manual labelling of millions of cars. Such a large-scale approach allowed them to come up with a complex cascade detector built on hand crafted Haar-like image representation. They report reaching 98% precision - 98% recall point. In this work, we aim to achieve similar performance, but without the prohibitively expensive human labelling.

2. Approach

The system is designed to work with all cameras independently. For every camera we consider the problem of computing the total number of cars that drive past a camera in a given time interval. In order to avoid over-counting, cars detected in multiple frames need to be identified as a single entity. Since the system is intended to work in the real-time, the processing speed for every frame is limited.

Optical systems for analyzing traffic generally detect and track cars from a video input. However, in our case, tracking-based approaches are not applicable because of very low framerate. Instead we combine cars detected in different frames in a probabilistic model that uses car appearance information and hints from scene geometry to identify same cars. Our approach includes four separate components: the background subtraction, the scene geometry, the car detector, and the probabilistic counting model (figure 2).

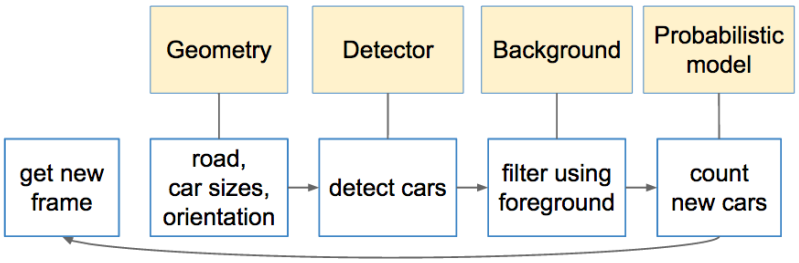

Fig. 2: system architecture

Understanding the scene geometry is the first part of the pipeline. Taking into account information about scene perspective helps to predict the expected size of a car at every particular pointof an image and the joint probability for this car to be detected in certain locations in two sequential frames. The scene geometry is first learned from clues like vanishing points and lane markings, and further refined with detected car trajectories.

After that, we detect vehicles on every new frame. We subtract the background based on a set of previous frames using Gaussian Mixture Models and use the foreground mask for detection. Separately, we train a Viola-Jones cascade detector. These two detectors complement each other. Replacing the simple Viola-Jones cascade model with the better Convolutional Neural Network detector is an ongoing work.

Finally, based on the car patches detected in the steps above, we find identical cars in pairs of sequential frames. Geometric constraints on location of a car in two frames and generated appearance-based features are used to build the transition probability matrix for every pair of frames. Car correspondences are then implied from this matrix.

3. Results

We built a system which provides car count in sparse and dense traffic (figure 3, left).

The performance is estimated on one video and reaches 86% recall - 95% precision point. The ongoing work for using Convolutional Neural Network (CNN) is expected to improve the performance and increase the variety of weather conditions where the system is applicable.

Fig, 3: left: example of detections; right: vehicles are generated artificially

Collecting training data for learning a CNN model is currently the main difficulty. Among the different techniques that facilitate data collection, the most promising one appears to be augmentatng the real dataset with artifical images. At the moment, thousand of training images can be generated automatically from 3D CAD car model collections (figure 3, right).