Technologies for Safe & Efficient Transportation

THE NATIONAL USDOT UNIVERSITY TRANSPORTATION CENTER FOR SAFETY

Carnegie Mellon University | UNIVERSITY of PENNSYLVANIA

Vehicle Detection and Counting at Road Intersections Using Video Data

FINAL RESEARCH REPORT

Camillo. J. Taylor

Contract No. DTRT-12-GUTC-11

DISCLAIMER

Project Title: Vehicle Detection and Counting at Road Intersections Using Video Data

Principal Investigator: Camillo. J. Taylor

Student Participants:
Takashi Furuya
Beau Kuhn
Miraj Shah
Amey Pendharkar

Collaborators:

Delaware Valley Regional Planning Commission
Scott Brady – Manager, Office of Traffic Monitoring
Paul Carafides – Traffic and Data Technician, Office of Traffic Monitoring

Objective:

Traffic analysis studies are commonly performed to inform policy makers about the rate of traffic flow and the prevailing traffic patterns at key intersections. In urban environments, like Philadelphia, one common approach to conducting these studies is by temporarily deploying a video camera that acquires footage over a period of study. This video is then analyzed to produce data on the routes that cars take at the intersection over that time period.

Figure 1: Still frame from a traffic video of a typical intersection

Once the video data has been obtained it can be manually annotated to obtain the required traffic count or turn information. As one can imagine, this is a time and labor intensive process. Alternatively, there are a number of commercial entities such as Miovision that have automated or semi-automated systems that can perform this function. Planning agencies can contract with these services for a fee to perform these counts.

What we are seeking to do in this project is to develop open source video analysis software that would reduce the cost of acquiring this data and, thus, make it possible to acquire more information about traffic flow in a timely manner. Ultimately we would like to develop algorithms that could be deployed on small, inexpensive embedded smart camera systems which could provide real time information about hundreds of intersections and give traffic planners and city operators an unprecedented level of information about traffic conditions.

Approach:

Over the duration of this project we have experimented with a number of approaches to tackling the video analysis problem. There are a number of factors that make this a challenging problem. Firstly in order to obtain accurate counts we need to segment vehicles from the roadway and from each other. Secondly, the system needs to accurately track each vehicle over time even as it stops, starts and turns. The non-nadir view of the traffic camera significantly complicates the problem. Firstly vehicles can and will occlude each other as they move through the scene. Secondly perspective effects cause the vehicles to grow and shrink in the image as they move through the scene. Finally shadows and distracting features are a major source of error which the software must contend with.

The basic architecture of our approach is outlined in the figure below:

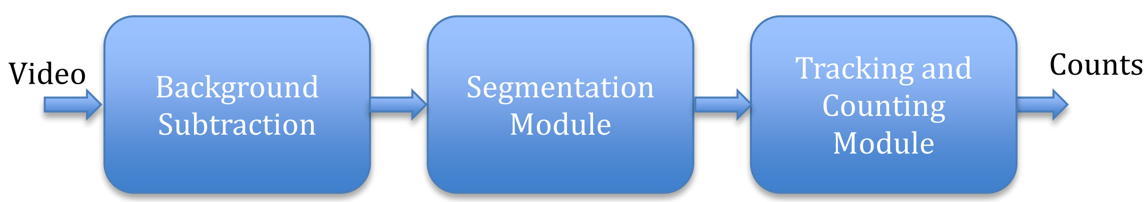

Figure 2: Block diagram of vehicle tracking architecture

The system is divided into 3 basic stages, a background subtraction system that identifies salient regions in the image, a segmentation module that is responsible for segmenting individual cars and the tracking and counting module that maintains a trajectory for each vehicle.

The background subtraction system is currently implemented by modeling the color distribution at each pixel as a unimodal Gaussian and then measuring the Mahalanobis distance between the current color vector and this background model. More sophisticated approaches using multi-modal Gaussians to better model shadows were investigated but these more expensive models did not appear to offer significant improvements in performance or shadow rejection and thus we are currently using the simpler model.

The background model is updated over time by choosing random samples from the last few images. Results obtained with this procedure are provided in the figure below.

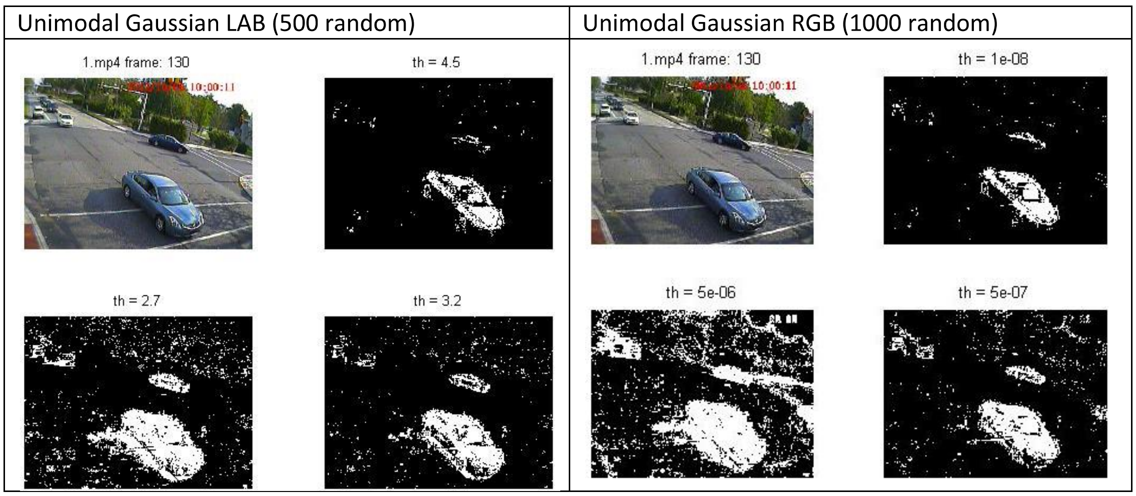

Figure 3: Output of background subtraction phase

The segmentation stage has proven to be the most challenging aspect of the vehicle tracking task. To date we have investigated two approaches to this problem. The first approach is a spatio-temporal analysis inspired in part by the pneumatic tube sensors that are commonly used for vehicle counts today. In this approach we specify a series of lines in the image that act as virtual pneumatic tubes and then record the extent to which each of these lines is occluded by the regions produced by the background subtraction phase. We then perform a spatio-temporal analysis of the signals associated with each track in search of patterns which correspond to vehicles driving in a certain direction over the image.

Figure 4: A series of virtual pneumatic tubes overlaid on the image.

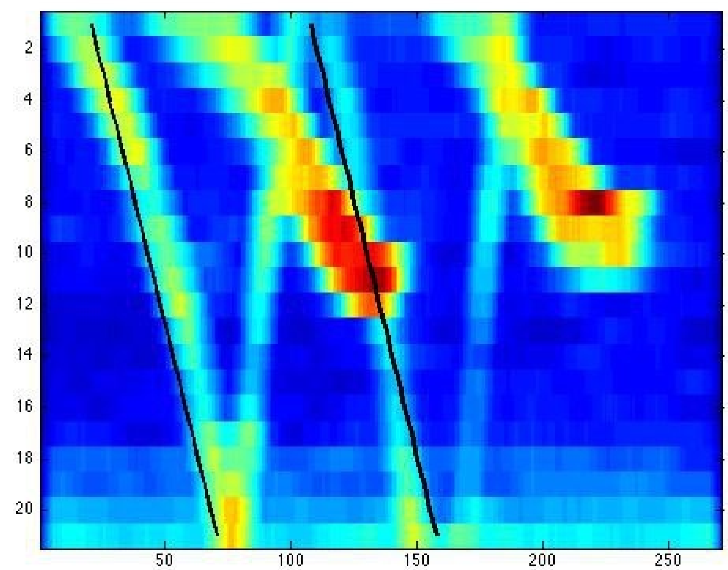

Figure 5: Results of a spatio-temporal analysis.

For example Figure 4 shows a series of lines drawn on the image while Figure 5 shows the occlusion level of each of the 20 lines over time for a segment of video. In this image we can delineate and count cars by finding appropriately oriented streaks in this spatio-temporal image. This approach is resilient to temporary occlusions and other distracting features but it requires the user to specify the possible trajectories that a vehicle could take through the intersection. Another issue that we encountered was that shadows associated with one vehicle could sometimes trigger spurious vehicle counts in an adjoining lane. These observations prompted us to investigate other approaches to the problem.

An approach that we are currently investigating proceeds by finding and tracking distinctive features in the video imagery. These feature trajectories can be obtained using the Lucas Kanade method or by detecting and tracking Harris corners. We then attempt to group the feature tracklets into coherently moving objects by exploiting the blob features extracted from the background subtraction process. The advantage of this approach is that it is insensitive to shadows since they do not induce coherent feature tracks.

The basic scheme is outlined in Figure 6. Here we see the features that are being tracked, the blobs that are detected by the background extraction procedure and the groupings that are inferred by the segmentation procedure. We are currently working on improving the segmentation scheme by enforcing the constraint that features that are grouped together should move coherently over time.

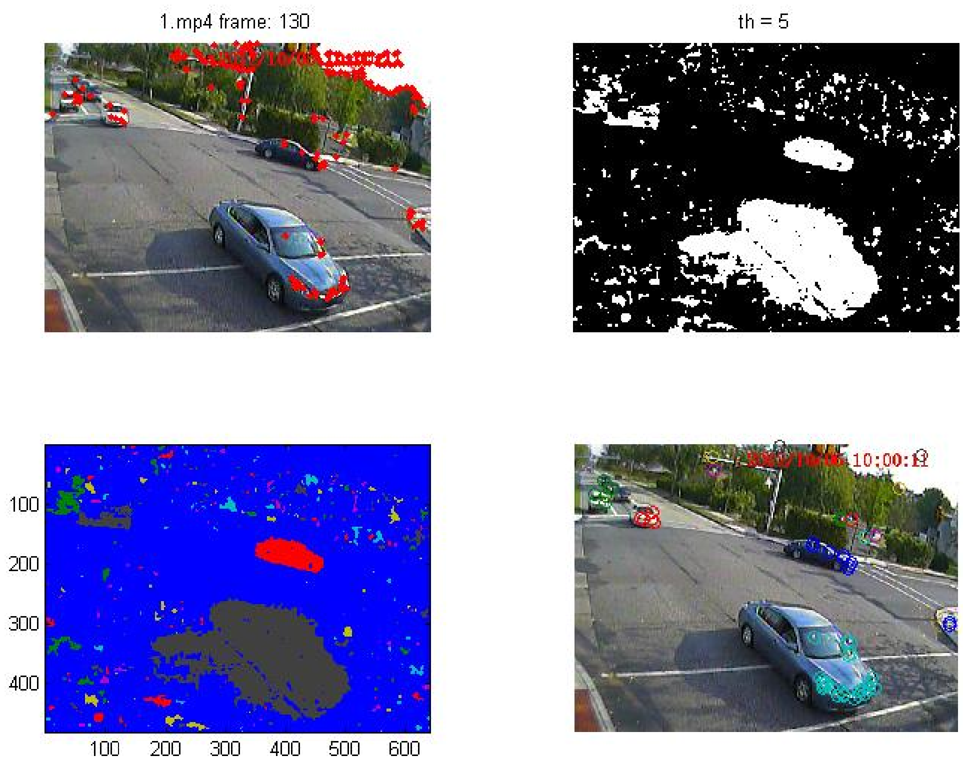

Figure 6: Feature based vehicle segmentation scheme showing how features are extracted and then grouped together.

Another idea that we have been leveraging in the course of this work is the thought that we can estimate the position and orientation of the camera relative to the roadway by exploiting a variety of cues like vanishing points and known or estimated road widths. We can use this information to effectively calibrate the camera which can then help us to reason about how the vehicles are moving on the ground plane and how they may be occluding each other. The figure below shows an example of how vanishing points associated with the roadway can be used to perform such a calibration.

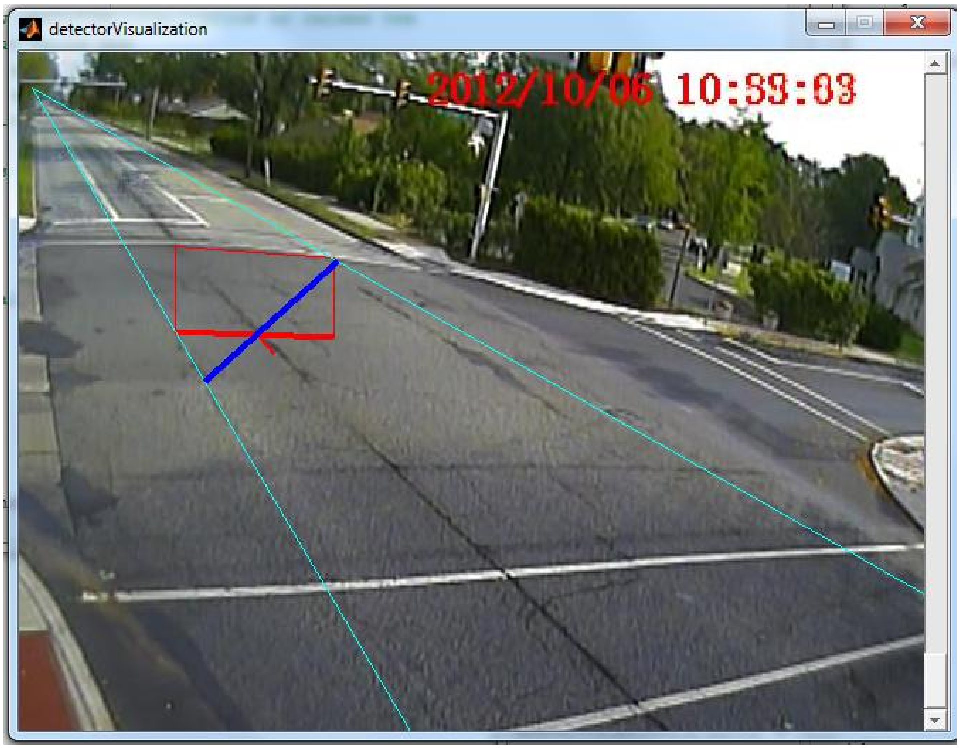

Figure 7: This figure shows our approach to determining the orientation of the camera wrt the intersection by exploiting vanishing points.

In summary, in our project we have been able to identify some promising approaches to video analysis that we hope will ultimately lead to a robust and effective car tracking and counting system. Our ultimate goal is to be able to demonstrate a system that is able to deliver accurate counts at 30 frames a second or better in a wide range of conditions. We are also hoping to develop an approach that could be implemented in real time on standard embedded processors like the ARM cores that are commonly used on camera equipped cell-phones. This would allow us to develop inexpensive embedded systems that could be widely deployed.