Unsupervised Framework for Traffic Counting: Speed Estimation Based on Camera Network Data

M.-W. Park[1], E. Palinginis[2], I. Brilakis[3], J. Laval[4], M. Hunter[5], and R. Guensler[6]

[1]Department of Civil and Environmental Engineering, Myongji University, 116 Myongji-ro, Cheoin-gu, Yongin, South Korea, 449-708; PH +82-31-330-6411; email: mwpark@mju.ac.kr

[2,4,5,6]School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, GA 30332; PH (404) 894-2278; email: [2]e_palinginis@gatech.edu, [4]jorge.laval@ce.gatech.edu, [5]michael.hunter@ce.gatech.edu, [6]randall.guensler@ce.gatech.edu

[3]Department of Engineering, University of Cambridge Trumpington Street, Cambridge, UK, CB2 1PZ; PH +44-1223-332718; email: ib340@cam.ac.uk

ABSTRACT

A variety of traffic data, such as traffic counts and speed estimations, can be harvested from camera network systems installed along highways. This is possible through computer vision-based traffic monitoring processes that are mainly composed of vehicle detection and tracking, and field of view calibration. Several such processes have been proposed, however they have not been fully validated on managing occlusion-based scenarios and generating reliable data over long periods of time and high volumes of traffic. This paper presents an effective; semi-automated method of detecting and tracking highway vehicles. The method integrates automated calibration of the field of view, detection and tracking. Trajectories, lanes, speeds and counts of tracked vehicles can be obtained from the videos using the proposed method. When a vehicle gets occluded by the other in adjacent lanes, the method identifies it based on the speed and acceleration, and terminates the tracking. When the vehicle reappears, it initiates a new tracking process. For validation, the framework is tested on videos recorded from CCTVs along the I-85 in GA, and evaluated on the accuracy of vehicle counting and speed. The tracked vehicles are counted when passing by pre-determined counting zones to avoid double counting. The speed results were compared with GPS data. The results indicate that the proposed system has a potential to minimize human intervention and provide reliable counting and speed data.

INTRODUCTION

Vision-based traffic monitoring processes including detection and tracking have been widely implemented to survey traffic conditions. Precise generation of traffic-related data based on vehicle-counting, however, is not possible with current technology given varying environmental conditions, such as illumination, and the

presence of occlusion. Most of the existing traffic counting systems still suffers from over-counting by detecting multiple smaller parts of larger vehicles, or under-counting due to the implementation of non-robust detectors. Incorrect detection is severely exaggerated under congested traffic, since many overlapped vehicles may be counted as one. In order to avoid over-counting, two mitigation strategies are employed to remove the previously detected tracking data and to solely consider the re-detected vehicle.

The processing was conducted on surveillance video data provided by the Georgia Navigator System in the State of Georgia, USA, which covers most of the highway corridor in metropolitan Atlanta. The proposed methodology is tested using various hour-long videos recorded from Close Circuit Television Cameras (CCTV) placed in various locations along the Georgia Highway Corridor. The first step relied on an automated calibration method based on the optimization of the extrinsic parameters of the system. To achieve the credibility of the input and the precision of the output, manual counts are employed as ground truth data for comparison with post-process analysis. The detection rate of feature-based detection as well as the removal rate of occluded vehicles is weighted individually. The results indicate the potential of the proposed technique to provide reliable traffic flow results based on accurate traffic counts. Though it relies on either the instantaneous or average speed, the method is effective because of its versatility in adapting various environmental variables such as camera view-angle.

BACKGROUND

Video based approaches to traffic monitoring have been actively investigated for decades. Road scenes recorded by a single camera and saved in digital video formats are processed to obtain traffic information including vehicle counts, speeds, and incident occurrences. Continuous research efforts have finally led to commercial ITS (Intelligent Transportation System) solutions for traffic surveillance system such as Autoscope (ImageSensing systems 2013) and Iteris (ITERIS 2013). The commercial solutions offer a full package capable of capturing and processing videos, and eliciting traffic condition information. While these solutions are increasingly implemented across the States, research efforts have continued to achieve higher accuracy and reliability. Image processing techniques used to extract traffic information from video data are classified into three main categories – detection, tracking, and calibration. Each individual vehicle is first recognized (detection phase), and its translational movement along the roadway is tracked (tracking phase). Calibration is typically required in order to map the image plane onto the real-world road plane.

For detection and tracking, various image features that can effectively set vehicles apart from other objects; or background, need to be selected. Background subtraction, HaarCascade, HOG (Histogram of Oriented Gradients), and point features are generally used and proven to be effective for tracking vehicles (Rodriguez and Garcia 2010; Feris et al. 2011; Bouttefroy et al. 2008; Kanhere and Birchfield 2008). Tracking involves one more step to infer the vehicle position based on its previous motion and appearance. Kalman filter, particle filter, and mean-shift

are the famous dynamic models employed for this position inference (Huang 2010; Scharcanski et al. 2011; Xiong et al. 2009). Particle filter was proven to be generally better than Kalman filter for nonlinear dynamic systems. Calibration is a mapping of between two 2D planes – image plane and road plane. A 3 by 3 homogeny matrix can represent the mapping. While the fundamental way of the homography determination is to use 4 or more point matches, several variations have been explored for roadside camera views (Kanhere and Birchfield 2010).

Recently, several research initiatives have been presented, reporting outperformance over the commercial solutions. Rodriguez and Garcia (2010) proposed an adaptive traffic monitoring system that performs well with various conditions in terms of illumination and traffic volume. Their system featured higher accuracy of vehicle count and speed than the commercial solutions. Also, Chintalacheruvu and Muthukumar (2012) also reported that their point feature based tracking system performs better than Autoscope with respect to vehicle speed. However, most of the existing methods have been tested on the videos that display a small number of lanes with zoomed-in views, and have not been tested on heavily congested road scenes. Accordingly, the methods are not fully validated for frequent and sever occlusion instances. Considering the field of view of the CCTV cameras on highways, severe occlusion has a critical impact on the traffic monitoring performance. The proposed framework is able to detect and track each vehicle independently, and effectively managing occluded vehicles.

METHODOLOGY

Field of View Calibration, Detection, and Tracking. A fixed camera view is calibrated to establish the transformation between the pixel coordinate and the real-world road coordinate. VWL method (Kanhere and Birchfield 2010) is selected for this purpose. VWL method requires three inputs – a vertical vanishing point (V), a known width (W), and a known length (L). Here, the directions of the width and length correspond to the transverse and longitudinal directions of the road, respectively. The three inputs are fed to the framework through manual interaction which enters width and length values and locates 6 points on the video frame. V is located the intersection of the two parallel side lines, W stands for the distance between the two lines, and L stands for the length of the middle line. The width and the length values are determined simply based on the standard dimensions of the lane width and the dividing line pattern.

The detection process consists of background subtraction and Haar-Cascade. Background subtraction locates the regions of moving objects. The main role of background subtraction is to constrain the image regions to search for the vehicle shape. The advantage of this step is twofold. It saves the processing time of Haar-Cascade and prevents from detecting false positives in the static background regions. The foreground blobs that result from background subtraction does not always contain a single vehicle. When congested, multiple vehicles generally appear merged into a single blob. In order to differentiate individual vehicles separately, shape feature based detection is applied to every foreground blob. Haar-Cascade is used for the shape detection, which is proven to be effective for low resolution views. Positive

training images are collected in an unsupervised way. Background subtraction applied to the highway scenes of light traffic conditions automatically generates cropped regions of a large number of vehicles that can be used directly for training.

The tracking algorithm of the proposed framework is based on the Ross et al.'s method (2008). The method infers on vehicle locations through particle filtering, employing eigen-images to model vehicle appearances. In the proposed system, instead of image coordinates, road coordinates are used as estimate variables in particle filtering. The x and y coordinates are modeled by independent Gaussian distributions, of which standard deviations are automatically adjusted to the vehicle's current speed. Additional constraint imposed to the particle filtering process is that y coordinates never decrease. Incrementally adjusted Gaussian distributions and the constraints on y coordinates allow better prediction of the vehicle location.

Automated Calibration. During calibration, each video sample has been calibrated to its unique characteristics using the Vanishing Point, Length and Width calibration strategy (Guiducci 2000). Considering that the width of each traffic lane is 12ft and the transverse length between each end of the tick mark of dotted lane 40ft, the intrinsic camera properties including focal length (*f*) or external variables such as tilt angle (φ) and pan angle (θ) can be computed. To automate the process, a total station has been used to specify the CCTV camera height. In particular camera height equals the same of the following: the height of the pole-base from the asphalt, the height of the pole from its base to the base of the mounted CCTV camera, and the height from the camera-base to the focal center.

The real world three-dimensional coordinates (x_r, y_r, z_r) are converted into an image coordinate (μ_i, v_i) in following format:

$$\begin{bmatrix}\mu_i\\ v_i\\ 1\end{bmatrix} = \begin{bmatrix} f & 0 & 0\\ 0 & -f\sin(\varphi) & f\,h\cos(\varphi)\\ 0 & \cos(\varphi) & h\sin(\varphi)\end{bmatrix}\begin{bmatrix}x_r\\ y_r\\ 1\end{bmatrix}, \qquad \text{(Eq. 1)}$$

assuming every point on the road is a planer object, which is not the case for vehicle objects. To incorporate this change, the transformation matrix is modified (eq.2) such to include the height of each vehicle. The height parameters are specified as initial height, h_o, final height, h_f, and increment value, h_c.:

$$\begin{bmatrix}\mu_i\\ v_i\\ 1\end{bmatrix} = \begin{bmatrix} f & 0 & 0\\ 0 & -f\sin(\varphi) & f\,(h-h_z)\cos(\varphi)\\ 0 & \cos(\varphi) & (h-h_z)\sin(\varphi)\end{bmatrix}\begin{bmatrix}x_i\\ y_i\\ 1\end{bmatrix}, \qquad \text{(Eq. 2)},$$

where the height is approximated as an array of 1xN size (eq. 3):

$$h_z \in [h_o, h_o+h_c, h_o+2*h_c, h_o+3*h_c, \ldots h_f], \qquad \text{(Eq. 3)}.$$

To automate the process, the output values for *f*, φ, θ, are estimated to fall within a range from -20 to +20 from the computed values; in increments of 2. Given the geometry of the entire calibration system and the method of maximum likelihood

estimation, the potential combination that maximizes the accuracy of the detected object gives the optimized calibrated output of f , φ, θ,. The definition of the method is given in equation 4.

$$\{(f\ ,\varphi,\theta,)mle\} \subseteq \left\{argmax\hat{l}(f\ ,\varphi,\theta|\mathrm{x}_1 y_1,\ldots,x_n y_n\ )\right\} \qquad \text{(Eq. 4)}$$

Occlusion Handling. Occlusion scenarios can be described by a combination of Clear state (denoted as C) and Occluded state (denoted as O). They include 1) a vehicle's clear view at first which then becomes occluded (C-O) 2) when a vehicle is initially occluded, but firstly appear once the occlusion is cleared (O-C), 3) when a vehicle's view is clear for the entirety of its movement (C), and 4) when a vehicle is completely occluded for the entirety of its movement (O). The occlusion type 4 is beyond the scope of this research as such occlusion is physically impossible to be recognized in any vision-based systems. For type 2 and 3, occlusion handling techniques are unnecessary as the vehicle will be detected and tracked once it grows in Clear state. However, type 1 can be problematic as it often makes wobble movements that affect the tracking accuracy. Therefore, the framework terminates the tracking of a vehicle in C-O transition by recognizing the wobble movement based on the following criteria:

$$S_F / S_{F-1} < S_{th} \quad \text{or} \quad S_{F-1} / S_F < S_{th} \qquad \text{(Eq. 5)}$$

$$S_F > S_{avg} / 4 \qquad \text{(Eq. 6)}$$

$$a_{avg,F} > a_{th} \qquad \text{(Eq. 7)}$$

where S_F is the scale of the tracked area at frame F, S_{avg} is the average scale value of all tracked vehicles, $a_{avg,F}$ is the average acceleration of a vehicle at frame F. S_{th} and a_{th} are threshold values which are set to 1.15 and 45 ft/sec^2. If a vehicle falls into the state satisfying both conditions of Eq. 5 and 6, or the condition of Eq. 7, then the object is likely to be occluded.

EXPERIMENT AND RESULTS

The proposed framework was implemented into a C# prototype and tested on videos taken from the I-85 highway in Atlanta. The performance of the framework was evaluated on vehicle counting and speed results.

Vehicle Counting. The proposed framework Fig. 1 illustrates a snapshot of the vehicle counting setup. Once a user specifies the detection zone, the counting zones are automatically placed on every lane next to the detection zone. When a detected and tracked vehicle passes the counting zone, it automatically adds a count and updates the counting on the top border. Both detection and tracking operates in the detection zone whereas only tracking functions outside the zone. The counting performance was analyzed by comparing with manual counting data. Two metric parameters were used for the analysis.

1) Correct Counting Rate (CCR) = counts that are vehicle / manual counting data

2) False Counting Rate (FCR) = counts that are non-vehicle / total counting data
3) Counting Error (CE) = 1 - total counting data / manual counting data

Table 1 summarizes the test results of four 15-minute videos. The framework performed well exhibiting over 95% CCR for Video 1, 3, and 4. It is inferred that the lower CCR of Video 2 was attributed to the foggy weather. The tests yielded low CE values as the missed vehicles were compensated by the false positives (non-vehicle).

Table 1. Vehicle counting results of four 15-minute videos

Video Sample	1	2	3	4
Manual count	3046	1550	2667	2669
# of lanes	7	5	7	7
vehicles/hr/lane	1740	1240	1524	1525
Traffic direction	away from camera	away from camera	toward camera	toward camera
CCR (%)	95.2	93.2	97.7	98.8
FCR (%)	4.9	5.3	2.1	2.0
CE (%)	-0.2	1.5	0.3	-0.8

Vehicle Speed. Vehicle speed is estimated based on the vehicle road coordinates and the video timestamps. Because of the timestamp noise and tracking error, vehicle speed data needs to be smoothed. In the proposed framework, the speed is smoothed by using moving average with 1-second interval. In the tests, GPS (Global Positioning System) was exploited to obtain ground truth data. Several floating tests were performed to collect both GPS and video data. A GPS unit was attached inside each test vehicle which was driven on the I-85 highway. Table 2 summarizes the speed estimation error of 6 floating tests, and Fig. 2 shows the speed comparison data for the floating test 6. In Table 2, the second row represents the overall average speed of the vehicles during the test. The speed error in the third row is calculated based on root mean square deviation.

Table 2. Speed estimation error

Test	1	2	3	4	5	6
# of speed data	586	316	345	288	111	293
Avg. speed (mph)	37.6	60.6	62.8	45.9	43.5	22.3
Error (mph)	1.99	1.56	1.14	3.24	0.95	1.54

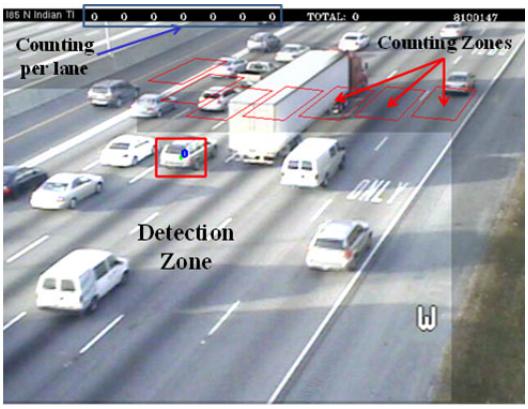

Figure 1. Vehicle counting setup

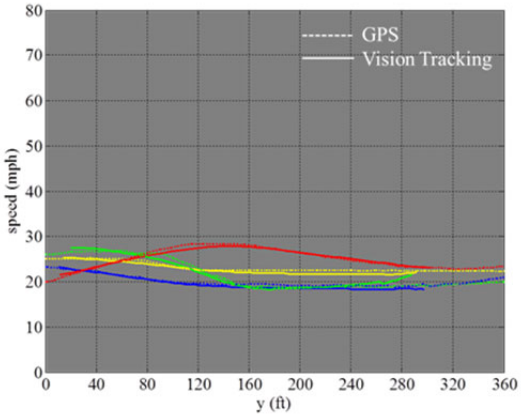

Figure 2. Vehicle speed comparison

CONCLUSION

In many states, video data of highway traffics are available from CCTV's located along the roads. Automated information extraction from the massive amount of video data will drive efficiency and deliver values to traffic monitoring practices. Aiming at making the best use of the videos, various vision-based solutions have been proposed. Although the previous solutions presented the ability to extract vehicle counts and speed data, they are not fully validated for the highway views that are seen very often from the roadside CCTV's. The view contains wide range of lanes, and occlusions by heavy trucks occur repeatedly. In this paper, the detection-tracking framework were developed and tested for feasibility in counting vehicles application, which were made possible by the optimization of the Haar-Cascade implementation. The framework provides further efficiency by automating the field of view calibration. The proposed framework is demonstrated to be effective as a reliable vehicle counting tool as a part of a vision-based traffic surveillance systems. Occluded vehicles are well managed in the counting system through unsupervised process of occlusion identification. Individual vehicle speeds are also accurately estimated by the framework. The systems also seem to exhibit reasonable robustness against changing illumination and object sizes. On the other hand, as general vision-based methods have, this method also has limitations to overcome. For example, it hardly works for night time views and severe weather conditions degrade its performance.

REFERENCE

Bouttefroy, P. L. M., Bouzerdoum, A., Phung, S. L., and Beghdadi, A. (2008). "Vehicle Tracking by non-Drifting Mean-shift using Projective Kalman Filter." Proc., the 11th International IEEE Conference on Intelligent Transportation Systems, 61-66.

Chaiyawatana, N., Uyyanonvara, B., Kondo, T., Dubey, P., and Hatori, Y. (2011) "Robust object detection on video surveillance." Proc., the 8th International Joint Conference on Computer Science and Software Engineering, 149-153.

Chintalacheruvu, N. and Muthukumar, V. (2012). “Video Based Vehicle Detection and Its Application in Intelligent transportation Systems.” Journal of Transportation Technologies, 305-314.

Feris, R., Petterson, J., Siddiquie, B., Brown, L., and Pankanti, S. (2011). “Large-scale vehicle detection in challenging urban surveillance environments.” 2011 IEEE Workshop on Applications of Computer Vision (WACV), 527-533.

Guiducci, A. (200). “Camera-Calibration for Road Applications.” Computer Vision and Image Understanding, 79(2), 250-266.

ImageSensing systems (2013). “Autoscope Video Detection.” <http://www.imagesensing.com/products/video-detection.html>

ITERIS (2013). “Abacus.” <http://www.iteris.com/products/software/abacus>

Huang, L., (2010). “Real-time multi-vehicle detection and sub-feature based tracking for traffic surveillance systems.” 2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR), 2, 324-328.

Kanhere, N. K., and Birchfield, S. T. (2008). "Real-time incremental segmentation and tracking of vehicles at low camera angles using stable features." IEEE Transactions on Intelligent Transportation System, 9(1), 148-160.

Rodriguez, T., and Garcia, N. (2010). “An adaptive, real-time, traffic monitoring system.” Machine Vision and Application, 21, 555-576.

Ross, D., Lim, J., Lin, R.-S., and Yang, M.-H. (2008). “Incremental learning for robust visual tracking.” International Journal of Computer Vision, 77(1), 125-141.

Scharcanski, J., de Oliveira, A. B., Cavalcanti, P. G., and Yari, Y. (2011). "A Particle-Filtering Approach for Vehicular Tracking Adaptive to Occlusions." IEEE Transactions on Vehicular Technology, 60(2), 381-389.

Xiong, C.-Z., Pang, Y.-G., Li, Z.-X., Liu, Y.-L., and Li, Y.-H. (2009). “Vehicle Tracking from Videos Based on Mean Shift Algorithm.” Proc., ICCTP 2009, 486-493.