VEHICLE DETECTION AND TRAFFIC DENSITY MONITORING FROM VERY HIGH RESOLUTION SATELLITE VIDEO DATA

George Kopsiaftis and Konstantinos Karantzalos

Remote Sensing Laboratory
National Technical University of Athens
Heroon Polytechniou 9, 15780 Zographos, Greece
gkopsiaf@survey.ntua.gr karank@central.ntua.gr

ABSTRACT

In this paper an automated vehicle detection and traffic density estimation algorithm has been developed and validated for very high resolution satellite video data. The algorithm is based on an adaptive background estimation procedure followed by a background subtraction at every video frame. The vehicle detection is performed through a further mathematical morphology and statistical analysis on the computed connected components. The traffic density has been estimated based on a lower resolution grid superimposed on the scene. In particular, at every subregion the number of the detected vehicles is calculated and the density is then estimated for the entire road network at every frame. The developed algorithm has been quantitatively evaluated. The quite promising results indicate the potentials of the proposed approach, while parallel gpu implementations can allow for real-time performance.

Index Terms— Skybox, earth observation, road network, transportation, traffic flow

1. INTRODUCTION

There is an increasing demand for optimal road network management based on efficient traffic monitoring in highly populated cities. To this end, several image-based vehicle detection techniques have been employed towards the development of automated monitoring methods for traffic flow detection and analysis along with the estimation of crucial parameters like density and speed.

Satellite-based methods have several advantages compared to the widely used video surveillance techniques from ground sensors (*e.g.,* CCTVs), since no sensor or camera installation and maintenance are required, while at the same time satellite imagery cover large areas, providing an integrated picture of the traffic conditions at a suburb or city spatial scale.

However, until recently, the spatial and the low temporal resolution of the available satellite data posed important constrains for efficient operational vehicle detection and traffic monitoring tasks. An important number of studies has been focused on the development of vehicle detection methods from single (static) very high resolution satellite data. In particular, Larsen et al. [1] employed an object-based maximum likelihood classification and shadow information procedure for Quickbird images. Gerhardinger et al. [2] evaluated pre-processing algorithms, while they digitized the road surfaces (road mask) to reduce the processed data. They experimented with Ikonos and Quickbird data of different atmospheric conditions. Moreover, Leitloff et al. [3], [4] focused primarily on the extraction of vehicle queues which were generated by the extraction of ribbons. Then, a search for single vehicles within the ribbons was performed using a least square optimization method. Last but not least, a morphological analysis as a pre-processing step followed by a neural network classification has been, also, proposed by Jin and Davis [5], [6] towards a binary separation of vehicle and non-vehicle pixels.

Moving from single image analysis to video sequence processing, in this paper, the goal was to exploit the recent very high resolution satellite video data that small satellite missions like Skybox Imaging Inc. can deliver. Towards this end, we have develop and validated an automated vehicle detection and traffic density estimation algorithm that can efficiently exploit the recent satellite video datasets.

2. METHODOLOGY

The developed methodology can be divided into three basic processing steps *i.e.,* background estimation, vehicle detection and traffic density estimation. During the first processing step the goal was to design an adequate and representative background estimation procedure able to capture all the non-vehicle static scene objects. However, note that since the satellite is moving uring the video acquisition the background is not actually stationary and is also moving. Several experiments with a different number of frames were performed. After a trial and error procedure the developed background estimation was based on an adaptive procedure during which the background was dynamically estimated based on the mean

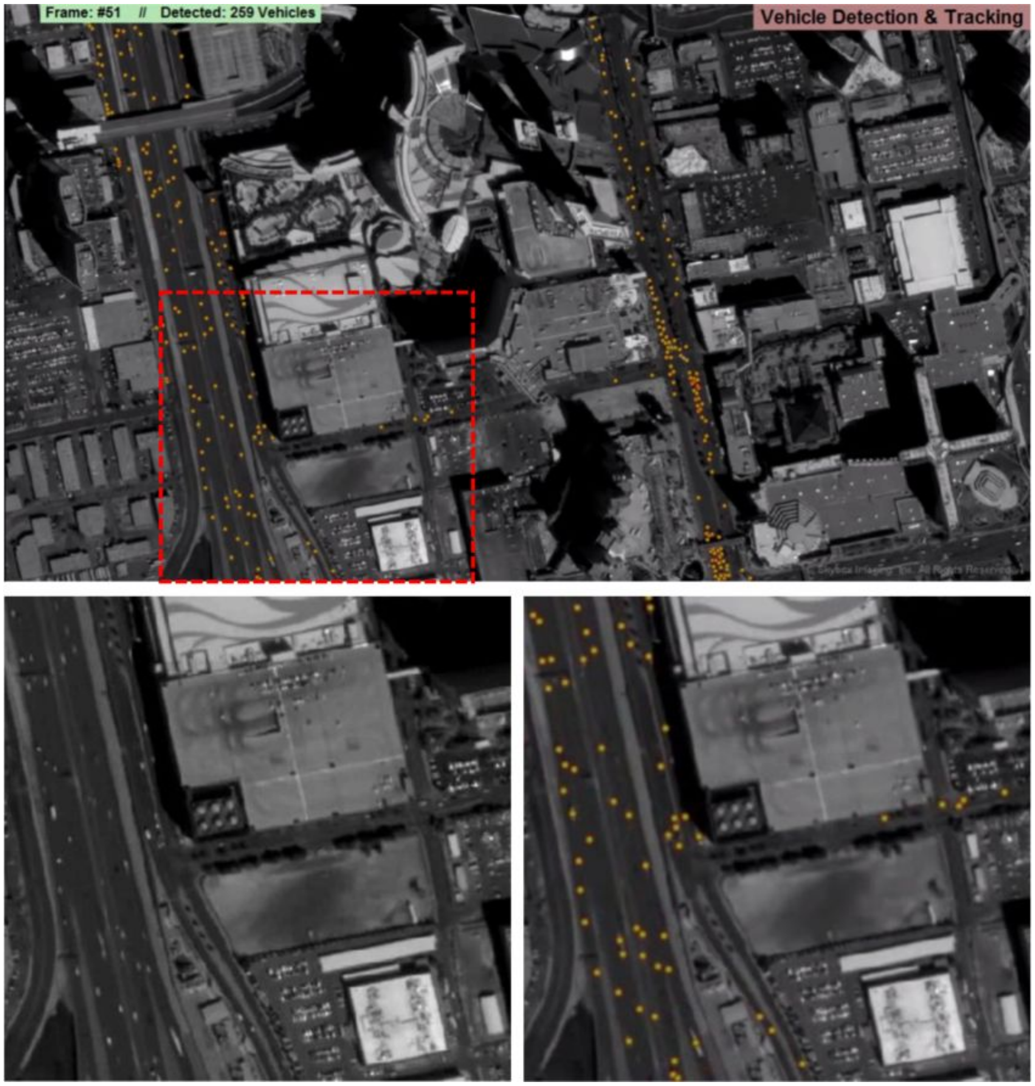

Fig. 1. Vehicle detection and tracking from satellite video data. The algorithm managed to detect *259* moving and non-moving vehicles (top, frame *#51*).

intensity value of approximately 300 frames.

The vehicle detection task was based on a background subtraction procedure followed by a mathematical morphology and statistical analysis on the resulted connected components. A global threshold value was computed using Otsu's method [7] after the subtraction. The traffic density has been estimated based on a lower resolution grid superimposed on the scene. In particular, at every subregion the number of the detected vehicles is calculated and the density is then estimated for the entire road network at every frame. In particular, the road network has been divided in a 50 by 50 meters grid, and the number of the detected vehicles is calculated at every grid cell. In all cases, prior knowledge concerning the exact road network boundaries has been integrated as in similar studies ([1], [2]).

Table 1. Skybox Imaging Satellite Video Dataset

Specifications	Sky-Sat-1 Video
Color	Panchromatic
GSD	1.1m at nadir
Duration	Up to 90s
Frame Rate	30 frame per second
Field of View	2km x 1.1km at nadir
File Format	H.264 (.mp4)

3. EXPERIMENTAL RESULTS AND VALIDATION

The developed methodology was applied on satellite video datasets. In particular, a full HD very high resolution video dataset acquired over Las Vegas from the SkySat-1 satellite on March 25, 2014 was employed. The video is provided by

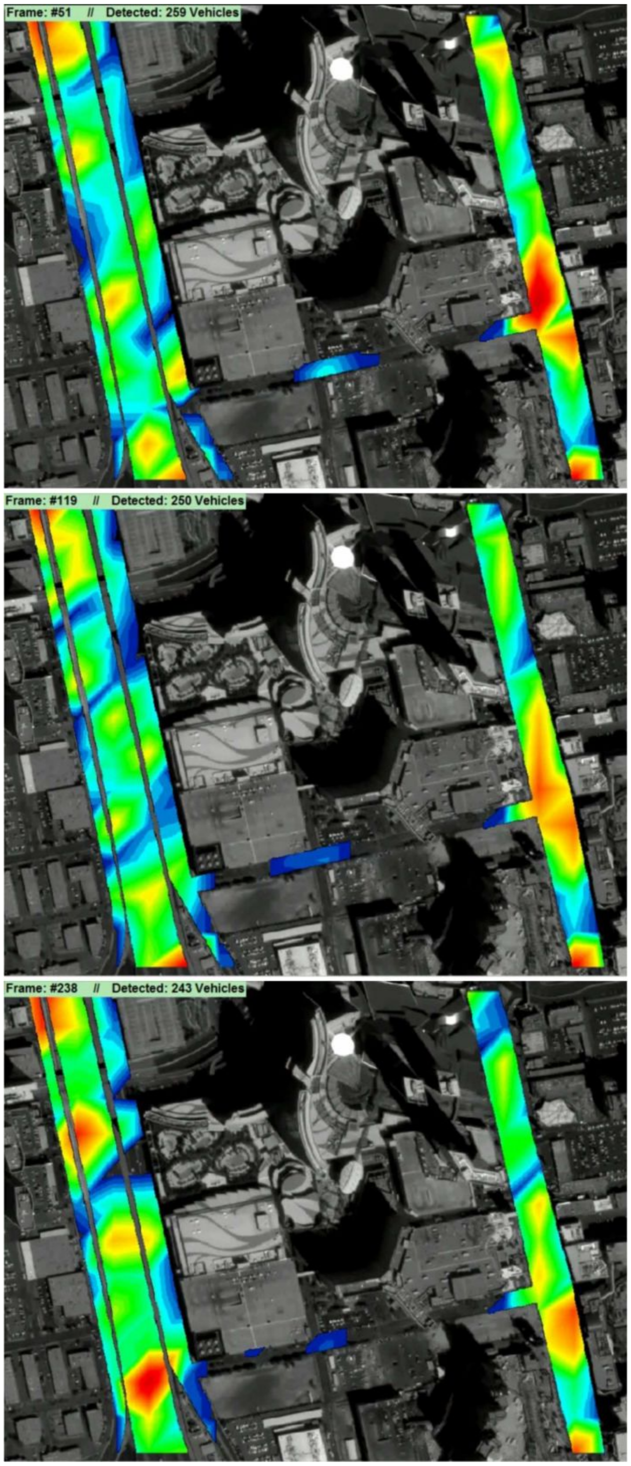

Fig. 2. Traffic density estimation from satellite video data. Results from three representative frames are shown along with the number of the detected vehicles at every frame.

Skybox Imaging and is publicly available[1]. The basic specifications[2] have been summed up in Table 1.

Table 2. Quantitative evaluation results across the entire video dataset. The maximum, minimum and mean values of completeness, correctness and quality indicators are reported.

	Completeness	Correctness	Quality
Max value	90%	84%	75%
Min value	80%	75%	65%
Mean	85%	80%	70%

Experimental results regarding the vehicle detection procedure for the entire dataset can be found here[3], while the estimated traffic density at every frame can be found here[4]. In Figure 1 results after the application of the developed algorithm are shown for frame *#51*. The algorithm managed to detect *259* moving and non-moving vehicles which are shown with a orange color superimposed onto the image. After a close look on the zoomed area (Figure 1, bottom) one can observe the efficiency of the developed methodology in detecting accurately the vast majority of scene vehicles.

In Figure 2 the estimated traffic density for three representative frames are presented as well. The estimated density highly correlates with the actual traffic flow, *e.g.,* increased values near the traffic lights, etc. In all frames the traffic density of the western road network was characterized by a relatively more stable spatial and temporal variability due to the fact that traffic flow was not interrupted by any traffic jam.

For the quantitative validation of the developed methodology, the standard quantitative measures of Completeness, Correctness and Overall quality have been employed based on equations 1-3. In particular, the True Positives (TPs) which are the vehicles that were correctly detected by the algorithm, the False Negatives (FNs) which are the vehicles that the algorithm did not managed to detect and the False Positives (FPs) which are the false alarms, were computed based on ground truth/ reference data. These ground truth data were derived after an intensive and laborious manual digitization of approximately 50 frames across the entire dataset.

$$\text{Correctness} = \text{TP}/(\text{TP} + \text{FP}) \tag{1}$$

$$\text{Completeness} = \text{TP}/(\text{TP} + \text{FN}) \tag{2}$$

$$\text{Quality} = \text{TP}/(\text{TP} + \text{FP} + \text{FN}) \tag{3}$$

In Table 2 results from the quantitative evaluation were summarised after the validation on the entire dataset. The maximum, minimum and mean values of completeness, correctness and quality indicators are shown. The detection completeness was in all cases more than 80%, while the Correctness higher that 75%. The mean values across the entire dataset were more than 80% both for Completeness and Cor-

[1] https://www.youtube.com/watch?v=lKNAY5ELUZY (accessed 20 November, 2014)

[2] http://www.euspaceimaging.com/products/143-skybox-prod-guide (accessed 10 May, 2015)

[3] https://www.youtube.com/watch?v=srJCy7H1gck

[4] https://www.youtube.com/watch?v=AH0HHt_Ya6M

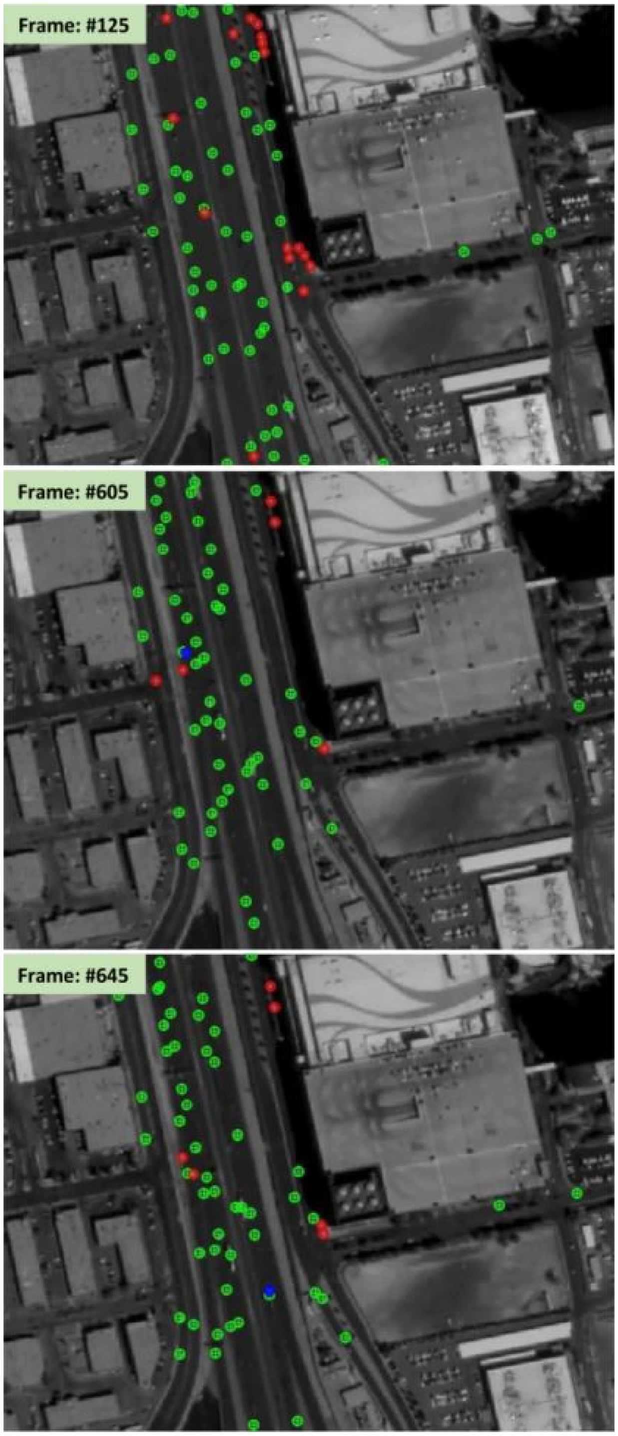

Fig. 3. Quantitative evaluation based on ground truth/ reference data. The computed TPs (with a green color), FPs (with red color) and FN (with a blue color) are shown for three representative frames.

rectness.

In Figure 3 three representative frames with the annotated TPs (with a green color), FPs (with red color) and FN (with a blue color), based on the reference data, are shown. In particular, one can observe that in frame #125 the number of FPs is relatively high especially near road margins. This in general is due to the fact that even slight camera movements between the sequential frames can fail to estimate efficiently the background in these regions resulting in certain FPs. In frames #605 and #645 the FPs are less but certain FNs are reported, also. These FNs cases occur mostly in locations where the continuous traffic flow has been interrupted *e.g.,* proximal to traffic lights where two or more cars are really close to each other and the algorithm detects them as one and not separated vehicles.

4. CONCLUSION

In this paper, a framework for traffic monitoring through vehicle detection and traffic density estimation from satellite video data has been introduced. The experimental results and the performed quantitative evaluation indicated that the developed methodology managed to deliver quite promising results. The overall quantitative validation reported completeness and correction detection rates over 80%. GPU implementations can, moreover, allow the real-time performance of the developed algorithm.

5. REFERENCES

[1] S.A. Larsen H., Koren, and Rune Solberg, "Traffic monitoring using very high resolution satellite imagery," *Photogrammetric Engineering & Remote Sensing*, vol. 75, pp. 859–869, July 2009.

[2] A. Gerhardinger, D. Ehrlich, and M. Pesaresi, "Vehicles detection from very high resulution satellite imagery," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Scienses*, August 2005, vol. XXXVI, Part 3/W24, pp. 83–88.

[3] J. Leitloff, S. Hinz, and U. Stilla, "Automatic vehicle detection in satellite images," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Scienses*, 2006, vol. 36, pp. 221–227.

[4] J. Leitloff, S. Hinz, and U. Stilla, "Vehicle detection in very high resolution satellite images of city areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 7, pp. 2795–2806, 2010.

[5] X. Jin and C.H. Davis, "Vector-guided vehicle detection from high-resolution satellite imagery," in *IEEE International Geoscience and Remote Sensing Symposium*, 2004, vol. 2, pp. 1095–1098.

[6] X. Jin and C.H. Davis, "Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks," *Image and Vision Computing*, vol. 25, no. 9, pp. 1422–1431, 2007.

[7] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.