Traffic Monitoring using Very High Resolution Satellite Imagery

Siri Øyen Larsen, Hans Koren, and Rune Solberg

Abstract

Very high resolution satellite images allow automated monitoring of road traffic conditions. Satellite surveillance has several obvious advantages over current methods, which consist of expensive single-point measurements made from pressure sensors, video surveillance, etc., in/or close to the road. The main limitation of using satellite surveillance is the time resolution; the continuously changing traffic situation must be deduced from a snapshot image. In cooperation with the Norwegian Road Authorities, we have developed an approach for detection of vehicles in QuickBird images. The algorithm consists of a segmentation step followed by object-based maximum likelihood classification. Additionally, we propose a new approach for prediction of vehicle shadows. The shadow information is used as a contextual feature in order to improve classification. The correct classification rate was 89 percent, excluding noise samples. The proposed method tends to underestimate the number of vehicles when compared to manual counts and in-road equipment counts.

Introduction

The monitoring of traffic conditions is necessary for development and maintenance of the road network. The current primary source of traffic statistics is measurement stations based on induction loops, which count vehicles that pass a given point in the road system over time. The most important information derived from these data is the Annual Average Day Traffic (AADT), a measure of the traffic distribution during a day at a specific point and averaged over a year. Nevertheless, this methodology has evident shortcomings due to the very limited geographical coverage of such a system. The necessary funding of covering the entire road network with such in-road counters is far from realistic, and alternatives are needed. Very high resolution satellite images, like QuickBird (0.6 m ground pixel resolution), may provide supplementary information to the traditional ground based sensors used for traffic monitoring. A satellite image covers large areas instantaneously providing a possible source of snapshot road traffic information. However, manual vehicle counting in images covering large areas would be a tremendous effort. A solution to this problem might be the use of an automatic image analysis methodology i.e., pattern recognition tailored to the detection of vehicles.

In cooperation with the local road authorities, the aim of our study was to compare in-road measurements with the traffic statistics that were estimated from automatic counts in QuickBird images. In this paper, we present the suggested approach for automated vehicle detection, consisting of segmentation followed by feature extraction and classification.

Extensive research has been performed on vehicle detection in aerial imagery, (e.g., Burlina *et al.*, 1997; Hinz, 2005; Schlosser *et al.*, 2003; Zhao and Nevatia, 2001). Recently, some research groups have also started to address vehicle detection in very high resolution satellite images.

Hinz *et al.* (2005) exploit contextual information to extract vehicle queues from QuickBird images of inner city roads. They also identify single vehicles using the width profile of the queues calculated from a gradient image. The approach is restricted to detection of vehicles in queues, and shows better results for dark vehicles than for bright vehicles. Gerhardinger *et al.* (2005) evaluates an inductive learning approach classification technique for vehicle detection on QuickBird and Ikonos images acquired over Baghdad, Iraq, with good results.

Jin and Davis (2004 and 2007) introduce a morphological preprocessing algorithm followed by a morphological shared-weight neural network classification of pixels into vehicle and non-vehicle targets. Ikonos imagery from three cities in Missouri is used for training and validation, with 85 percent detection rates and few false detections. The training data is derived from manual delineation of vehicles in the images. Zheng *et al.* (2006) use a similar approach. Zheng and Li (2007) also use morphological transformation of the image to identify candidate vehicles, but then an artificial immune system approach is used to create a set of templates called antibodies for vehicle detection. The experimental data in the two latter cases consist of QuickBird images of city scenes, including roads and parking lots.

Sharma *et al.* (2006) present three different approaches for vehicle detection in Ikonos images. Of these, the best performance is obtained using a pixel-based Bayesian Background Transformation approach, which requires the existence of a high quality estimate of a background image.

Alba-Flores (2007) detects vehicles in Ikonos images of U.S. one-way highway segments using two different thresholding approaches; a multiple threshold approach, and Otsu's method for bright and dark objects separately. These methods are tested on selected sections of the imaged highways that must adhere to severe restrictions.

A recent study performed by Pesaresi *et al.* (2008) addresses estimation of vehicle velocity and direction, exploiting the fact that there is a time lag between the acquisition of

Norwegian Computing Center, P.O. Box 114, Blindern 0314 Oslo, Norway (siri.oyen.larsen@nr.no).

Photogrammetric Engineering & Remote Sensing
Vol. 75, No. 7, July 2009, pp. 859–869.

0099-1112/09/7507–0859/$3.00/0
© 2009 American Society for Photogrammetry and Remote Sensing

the panchromatic and multi-spectral images. The vehicles are found by rough manual selection of target points, followed by automatic refinement of the location of the targets based on morphology.

Common to all the previous research that we have seen is that the methods have been developed for American highways or roads in large urban environments around the world. Our attempt is focusing on the difficulties related to analysis under the conditions present at Norwegian roads. Norway is a sparsely populated country, and roads are typically narrow and often close to a forest on one or both sides. A frequently encountered problem is that much of the road is hidden by tree shadows. This problem is worst at wide view angles from the satellite. Furthermore, the traffic density is very low compared to published studies from highways in other countries.

The work presented in our paper is building upon the results of the ESA (European Space Agency) project "Road Traffic Snapshot" which addresses a possible future service for counting vehicles in satellite images and generating traffic information based on these counts (Aurdal *et al.*, 2007; Eikvil *et al.*, 2008). Some of the methods presented here are advanced versions of the methods presented in Eikvil *et al.* (2008). However, while their study site is the City of Oslo, we attempt to take a further step towards developing a robust vehicle detection method that handles a wider range of imaging and traffic conditions as well as road types than previous approaches.

Methods

Experimental Data

To be able to detect vehicles, satellite images of high resolution are required. We have chosen the QuickBird satellite with 0.6 m ground resolution in the panchromatic band. The data consists of five satellite images from the QuickBird image archive, and covers different parts of Norway for the period between 2002 and 2006. Table 1 presents an overview of the data, including image size, length of the extracted roads, and average traffic density. All the images are acquired in the summer season, i.e., without snow covering the roads and with enough sunlight to be able to detect vehicles. The selection of image data for our study was made so that it represents a variation of viewing and lighting conditions, i.e., the combination of road direction, sun elevation, and view angle, which has an important effect on how objects and shadows appear in the image. The extracted roads represent typical Norwegian traffic conditions; narrow roads, often close to a forest on one or both sides, and low traffic density compared to other countries. To give an impression of the given conditions, a section of the Sollihøgda image is shown in Figure 1; the length of the road segment in the illustrated area is 2.2 km. Traffic counts from in-road equipment measured at the same times for the corresponding roads were provided by the Norwegian Public Roads Administration and were used for comparison with the image analysis based counts.

Preprocessing

Road masks are applied to the images in order to restrict vehicle detection to roads. Thus, ideally, all pixels outside the road are considered to be background pixels. As a first approach, the road masks were generated automatically, using vector data representing the mid-line of the road accompanied with a parameter representing the width of the road. A buffer mask was grown around the mid-line vectors. However, the road width was not sampled densely enough to yield an adequate representation of the road, and the maximum width had to be applied to the entire road to ensure not losing parts of the road surface. Furthermore, rectification by manually selected reference points was necessary. To avoid resampling of the image, the road mask was rectified to match the image. For most of the images, the match between the rectified mask and the image was not satisfactory. Therefore, we make road masks by manual digitalization. Gerhardinger *et al.* (2005) report on similar problems using a vector layer representing road centerlines and a standard buffer width throughout the road network. Their good classification results were dependent on precise vector data describing the road surfaces, and in the absence of automatic procedures for extraction of road surface, they derive road masks manually by digitizing on screen.

In addition to the road mask, a vegetation mask is applied. Vegetation that may obstruct the view of parts of the road includes the crowns of trees by the road side and plants growing in-between different lanes. The vegetation mask was derived from the multispectral information from the same scene. The normalized difference vegetation index (NDVI) is computed from the multispectral image after resampling to the resolution of the panchromatic image using cubic interpolation. We then find the appropriate threshold from the application of Otsu's algorithm (Otsu, 1979) to the resulting NDVI image, and use this to produce a vegetation mask. In the end, the road and vegetation masks are combined to produce a masked panchromatic image, i.e., a panchromatic image where all vegetation and/or non-road pixels are set to black (Figure 2).

Segmentation

The segmentation routine is based on Otsu's method for unsupervised threshold selection (Otsu, 1979). Given a gray tone image with gray levels $1, \ldots, L$, Otsu's method divides the pixels into two classes with gray levels $0, \ldots, k$

TABLE 1. EXPERIMENTAL DATA. THE TRAFFIC DENSITY GIVES THE AVERAGE NUMBER OF VEHICLES PER KM AT THE TIME OF IMAGE ACQUISITION (CALCULATED BASED ON MANUAL COUNTS IN SOME OF THE EXTRACTED ROAD SEGMENTS)

Location	Image Area (km^2)	Road	Length of Road Mask (km)	Traffic Density (vhcls/km)	Date (mm.dd.yy)	Time (UTC)	Upper Left Latitude	Mean Sun Elevation Angle	Mean Off nadir View Angle
Bodø	32	RV80	9	–	07.21.03	10:32	67	43	4.4
Kristiansund	29	RV70 (inner city)	5,6	20,9	06.19.04	10:56	63	50	7.9
		RV70 (outer city)	5,8	5,5					
Østerdalen	59	RV3	31,4	1,4	08.10.04	10:39	62	43	7.3
Eiker	154	EV134	12,2	–	06.07.02	10:42	60	53	12.9
		RV35	20,5	7,3					
Sollihøgda	52	EV16	26,6	6,7	05.10.02	10:32	60	47	12.5

Figure 1. One part of the Sollihøgda image. The white lines represent the borders of the road mask. In this sub-image, 19 vehicles were observed, and the road segment is approximately 2.2 km.

and $k+1, \ldots, L$, respectively. The threshold k is determined by computing the gray level histogram and selecting the gray level that maximizes the between-class variance. In the histogram of the masked panchromatic image (Figure 3) the road (asphalt) pixels constitute the dominating mode. (Background pixels are not included in the histogram). Dark colored vehicles and other dark segments are represented by a smaller peak in the histogram. The long tail in the high intensity part of the histogram corresponds to the class of bright segments on the road, appearing in a wide range of intensities. Note that Otsu's method may easily be extended to the multi-class problem, e.g., finding two optimal thresholds simultaneously. However, experiments showed that using the three-class version of Otsu did not provide two

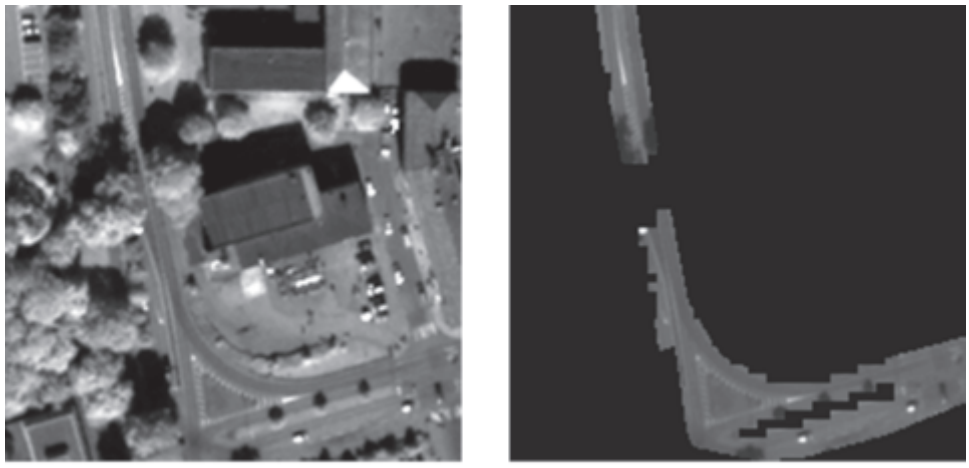

Figure 2. Vegetation covering parts of the road: (a) Panchromatic image, and (b) masked (road and vegetation) panchromatic image.

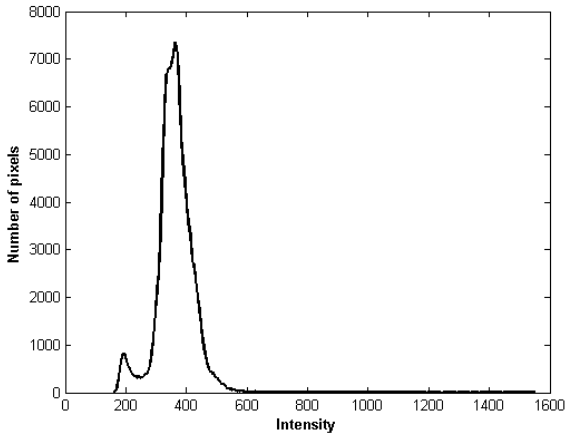

Figure 3. Image histogram.

thresholds that are able to make the desired separation between the classes in our application. One reason is perhaps the fact that the division line between the road class and the bright segment class is very diffuse (the histogram does not appear to be trimodal).

The panchromatic image is thresholded in two stages, for dark and bright segments, respectively. Selection of the right threshold is critical for obtaining a successful segmentation; using a too strict threshold yields poorly defined or fragmented segments, while less strict thresholds tend to produce too many segments and include many unwanted objects, such as road marks (see Figure 4). To overcome this problem, we perform so-called hysteresis thresholding. This is traditionally an edge detection method, where the input image is thresholded with a *higher* and a *lower* threshold, producing so-called *strong* and *weak* edges. All the strong edges are kept in the output image, while a weak edge is only kept if it is connected to a strong edge. This method gives well-connected edges while eliminating isolated, noisy edges. Similarly, we select a strict and a less strict threshold at each stage (for segmentation of bright objects, the stricter threshold is the higher one, while for segmentation of dark objects, the stricter threshold is the lower one), and keep segments resulting from the least strict threshold only if it contains a segment resulting from the stricter threshold. Each threshold is found by restricting the focus to a subset of the image histogram. Let I_{min} and I_{max} denote the minimum and maximum intensity values, greater than zero, that are present in the image, and let μ and σ denote the mean and standard deviation, respectively. We then define the thresholds as follows:

Segmentation of dark objects:

- low threshold found by Otsu applied to the histogram on the interval $[I_{min}, \mu - \sigma]$
- high threshold found by Otsu applied to the histogram on the interval $[I_{min}, \mu - 0.5\sigma]$.

Segmentation of bright objects:

- low threshold found by Otsu applied to the histogram on the interval $[\mu + \sigma, I_{max}]$
- high threshold is $\mu + 3\sigma$.

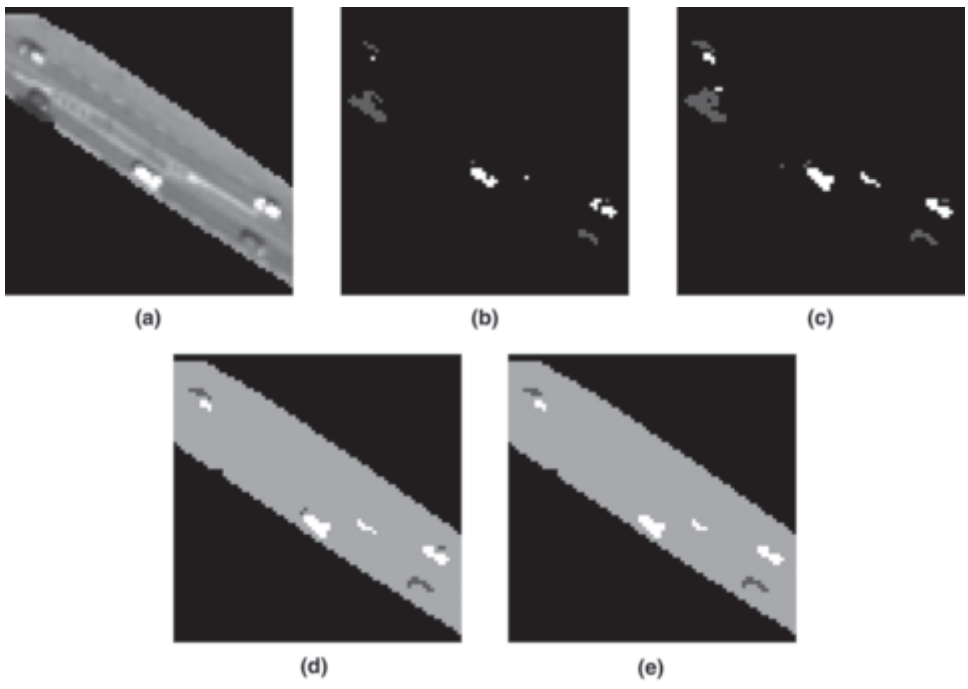

Figure 4. Segmentation steps: (a) masked panchromatic image, (b) result of stricter dark and bright thresholds (dark threshold yield gray segments; bright threshold yield white segments), (c) result of less strict dark and bright thresholds (graytones as in (b)), (d) combined segmentation result (dark segments in darker gray, bright segments in white, road in lighter gray), and (e) segments after pre-classification (gray tones as in (c)).

Feature Extraction

We have examined various features that describe the segments. The features are used to discriminate between different types of objects, i.e., vehicles from other objects. The following features were selected:

- mean intensity of the object
- mean gradient of the object (using the Sobel operator)
- standard deviation of the intensity within the object
- object length (defined as the length of the bounding box)
- the object's 1st Hu moment (Maitra, 1979)
- the spatial spread of the object (calculated using normalized central moments of the object)
- object area
- object elongation
- distance to nearest vehicle shadow.

The first six features are used for statistical classification by maximum likelihood. They were selected from a larger set of features using statistical feature selection methods. The object area and elongation are used in a rule-based pre-classification step, where obvious non-vehicle segments are discarded. The object area must neither be too large nor too small. We also require that the elongation adheres to a given interval. Furthermore, we assume that the mean intensity value within the object is above or below preset thresholds, and the object gradient must also obey a minimum threshold value. Finally, dark segments that overlap the edge of the total mask (i.e., the road and vegetation masks combined) are also rejected during pre-classification. This is done because tree (or building) shadows are frequently observed at the edge of the road (Figure 5). Finally, the distance to nearest vehicle shadow is used as a post-classification step, as described later.

Figure 5. A common sight on Norwegian highways: shadows cast from trees. The white contour delineates the edge of the road. Dark segments overlapping this edge should be discarded.

Prediction of Vehicle Shadows

Many of the segments represent vehicles shadows. These are of concern to us, because we need to separate them from the vehicles during classification. Most of the bright vehicles cast clearly visible vehicle shadows, while it is visually impossible to distinguish a dark vehicle from its shadow (Figure 6). We have developed a new approach regarding the handling of vehicle shadows. Our approach is based on using information about the direction and elevation of the sun at the time of image acquisition to predict shadows. This idea is not new; Hinz (2005) carries out an accurate computation of a car's shadow projection based on date, daytime, and image orientation parameters. However, the manner in which the predicted shadows are used for vehicle detection differs between Hinz' approach and ours. Hinz is studying vehicle detection in high-resolution aerial imagery, and incorporates the predicted shadow into an explicit 3D vehicle model, which is then used in a matching algorithm. On the other hand, we predict a vehicle shadow mask, i.e., an image containing segments that represent vehicle shadows. From

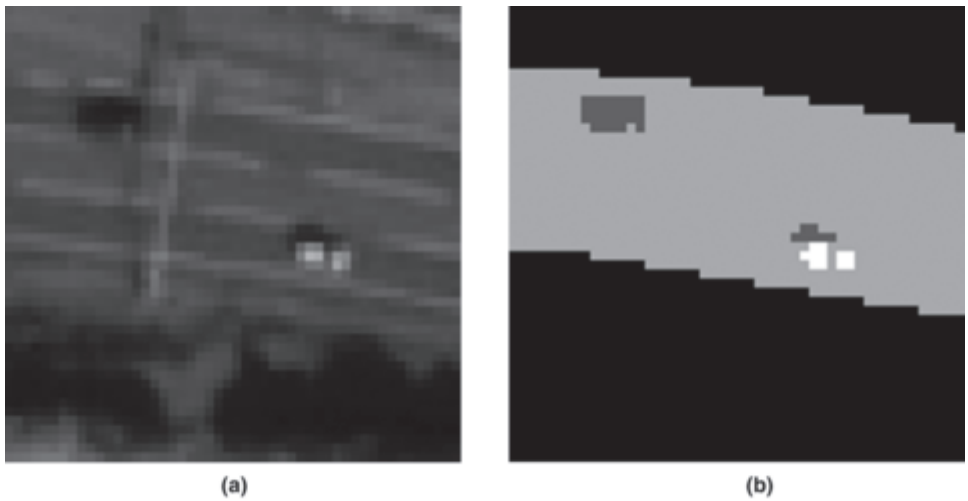

Figure 6. A bright and a dark vehicle in (a) panchromatic image and (b) segment image. Note that we only see the vehicle shadow of bright vehicles. Also note that the bright vehicle is fragmented into two bright segments.

this mask, we estimate the *distance to nearest shadow* as one of the features describing the object, i.e., during the feature extraction step. This feature is then used for classification, which is described in a separate section below.

The calculation of a shadow mask is carried out as follows. Based on information about the sun direction, we first calculate an image mask that indicates the expected shadow zones of the vehicles that are found in the image. The expected shadow zone of a bright object is a sector lying directly opposite the object with respect to the direction of incoming sunlight. The expected shadow zones of all the bright objects in the image are found by dilation of a segmented image, representing bright objects only (Figure 7b), with a structuring element representing the expected shadow sector. The direction and length of the shadow sector are determined from the two parameters sun azimuth and sun elevation, respectively. These two parameters may vary between different images. We use four different directions to estimate the direction of the shadow zone, and we use a 90° wide zone, pointing north, east, south, or west. For example, if the sun azimuth is between 135° and 225°, the sun enters the image scene from the south, and the expected shadow zone lies north of the objects, so this is where to look for segmented dark objects that could be shadows. The length of the shadow is given by the average vehicle height divided by the tangent of the sun elevation angle. For QuickBird images an average vehicle height of 3 pixels = 1.8 meters is used. As an example, the shadow structure element in Figure 7d would be used if the expected shadow zone was *four* pixels (= 2.4m) long and located *north* of the vehicle. Dilation of image I with image (structure element) S is defined as

$$I \oplus S = \{z \mid \hat{S}z \bigcap I \neq \varnothing\},$$

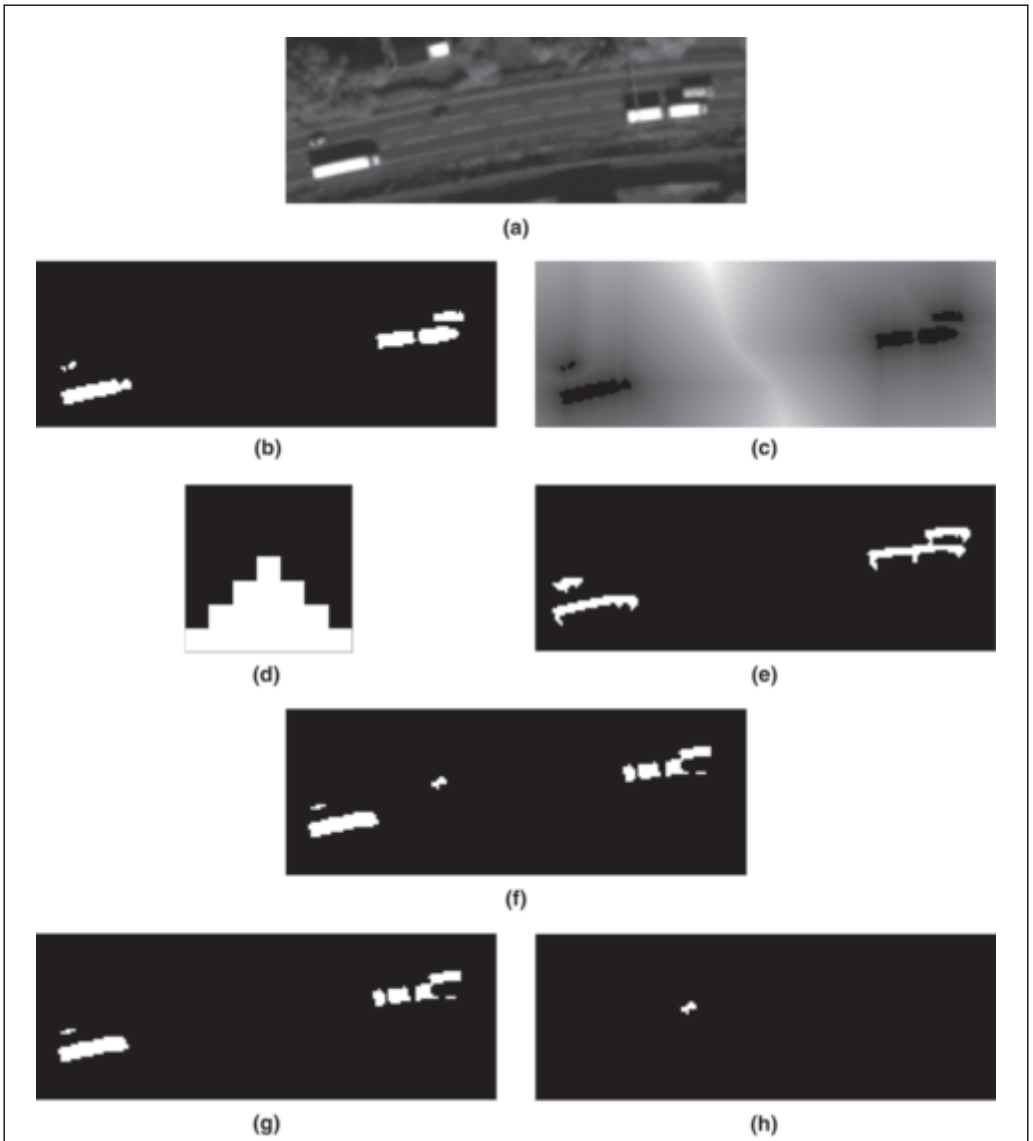

Figure 7. The entire vehicle shadow prediction process. The bright segments (b) are dilated with the shape of the expected shadow zone (d), and the segments are then subtracted from the dilation result, yielding an image of the expected shadow zones only (e). Together with the distance map of bright segments (c), this decides which dark segments (f) are vehicle shadows (g), and which are not (h).

where $\hat{S}$ denotes reflection of S, defined as:

$$\hat{S} = \{x \mid x = -s, \text{for } s \in S\},$$

and S_z denotes translation of S by (a vector) z (Gonzales and Woods, 1992; page 519). The dilation result represents an image of the bright objects together with their expected shadow zones. The bright object segment image is subtracted from the result of dilation, yielding an image representing the expected shadow zones only (Figure 7e). This image is then compared with a segmented image of dark objects that lie close to a bright object. Wherever there is overlap between these two images, the dark object is assumed to be a shadow. The segmented image of dark objects that lie close to a bright object was first found by calculating the distance map to the bright segment image (Figure 7c), and then combining this information with the dark segment image (Figure 7f), keeping only the dark segments whose distance to a bright segment is below a preset threshold (Figure 7g). The result of this process is a shadow mask. This mask is used to estimate the distance to the nearest shadow, which is a feature that is calculated for each object resulting from segmentation. Ideally, the shadow distance is zero for objects that are true vehicle shadows, positive for other dark objects, small for bright vehicles, and (relatively) large for road marks and other bright segments that may be confused with vehicles.

Classification

We use the well-known maximum likelihood method for classification of segments into one of six classes:

1. Bright car
2. Dark car
3. Bright truck
4. Bright vehicle fragment
5. Vehicle shadow
6. Road mark (arrow shaped).

The motivation behind the bright vehicle fragment class is to include segments that belong to a bright vehicle, but fail to capture the whole object. Bright colored cars are frequently seen as two smaller fragments, probably due to the dark interior of the car being visible through the front or rear window (Figure 6). In other cases, the bright car is represented by one fragment only, usually due to bad contrast which gives rise to a suboptimal segmentation. Bright trucks have higher mean intensity and are larger than cars, thus a separate class for large sized vehicles is defined. In our experimental data, there are only a couple of examples of heavy dark vehicles, hence no dark truck class is defined. Two rather homogenous groups of non-vehicle objects are observed on such a frequent basis that they have been given a dedicated class of their own; heads of bright arrows painted on the road (directions for traffic), and vehicle shadows. The remaining objects do not fall into one specific class of shape or intensity. These segments are defined to be noise objects and include road marks of various geometric shape (other than arrow), (parts of) bridges across the road, road signs by the roadside or across the road, or shadows of these, (parts of) roundabouts, erroneously segmented vehicle objects, grown together in pairs, with road marks, or other objects (Figure 8).

We assume that the feature vectors within each class are normally distributed, and that we have general covariance matrices (correlated features). Furthermore, we assume that the prior probability $P(\omega_\kappa)$ of belonging to class κ is $P(\omega_\kappa) = N_\kappa / N$, where N_κ is the number of training samples from class κ, and N is the number of training samples from all the classes in total. Finally, we made experiments where we introduced the possibility of rejecting a segment if none of the class posterior probabilities were reasonably high. The purpose of such an option is to exclude objects that do not belong to any of the defined classes. However, we were not able to see any difference in maximum probability for such outlier objects, hence, we did not use this option in the final algorithm.

Vehicle shadows are often confused with dark vehicles and vice versa, even after application of the maximum likelihood classifier. We therefore seek to reduce the number of misclassifications between vehicle shadows by revising the segments that are classified into one of these classes. The post-classification is based on the distance to nearest shadow feature, calculated from the vehicle shadow mask, as described above. Specifically, dark segments that are included in the shadow mask will have zero shadow distance, hence the post-classification gives all dark segments with zero shadow distance the label of the shadow class. The shadow distance information is also used to improve the classification of bright vehicle fragments into road marks. These two classes share similar shape and intensity features. However, while vehicle fragments often cast a detectable shadow, road markings do not. The classification of a road marking is changed to

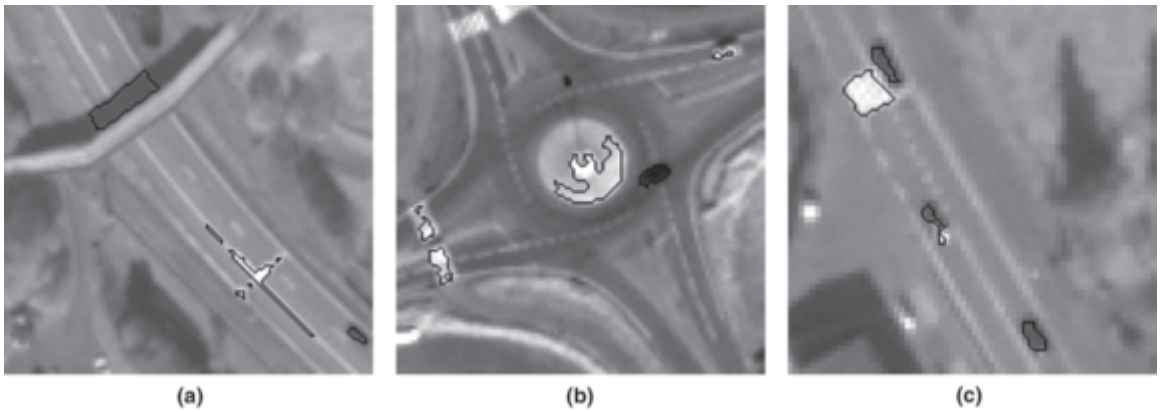

Figure 8. (a) through (c): Examples of noise segments, i.e., segments not belonging to any of the defined classes.

Figure 9. Final detection results. Detections are enclosed in white squares. The image shows part of the inner city road in the Kristiansund image.

bright vehicle fragment if its distance to a shadow is less than a (small) preset threshold given in pixel units. Final detection results are illustrated in Figure 9.

Experiments

The main classification was applied to all segments characterized as potential vehicles by the pre-classification routine. The parameters for the normal distributions were estimated using a training data set, consisting of sub-images of satellite scenes from Bodø, Kristiansund, Eiker, and Sollihøgda (see Table 1). The segments were labeled manually, using labels 1 through 6 for the respective classes. As some of the segments represent objects of low contrast or cluttered objects, labeling is not necessarily a straightforward task. Aurdal *et al.* (2007) reported 83 percent agreement between two individual, manual classifications. However, their experiments were performed on images of inner city roads, where cars appear closer to each other, and one expects more clutter. Our experiments are mainly conducted on highways with less traffic, although some city roads are included. A total of 787 samples were used for training, of which there were 123 belonging to class 1, 152 to class 2, 37 to class 3, 152 to class 4, 206 to class 5, and 117 to class 6. Only segments belonging to one of the six classes were included in the training set. For testing, we used a different set of sub-images of satellite scenes from Kristiansund, Eiker, Sollihøgda, and Østerdalen, which altogether gave 372 test samples. The test set includes 73 noise samples (segments not belonging to any of the defined classes), because it is interesting to see whether these samples are classified into vehicles or one of the non-vehicle classes. The fraction of noise segments compared to the total number of segments was 2 percent, 20 percent, 22 percent, and 30 percent, for Eiker, Sollihøgda, Kristiansund, and Østerdalen, respectively.

The classification results are shown in Table 2. The correct classification rate, i.e., the fraction of correctly classified segments, is 89 percent, excluding noise segments, i.e., segments not belonging to any class, as defined above. The confusion matrix (Table 2) shows that 12 out of 139 bright vehicles are classified as road marks. These are typically small vehicle fragments with low mean intensity compared to other bright vehicles. We also see that 10 out of 77 dark vehicles are classified as vehicle shadows. The vehicle omission error for the classification, i.e., the fraction of vehicles that are omitted in the classification compared to all the vehicles in the classification, is 10 percent. On the other hand, the vehicle commission error, i.e., the fraction of non-vehicle segments that are classified as vehicles compared to the total number of vehicles in the classification, is 5 percent. However, counting noise samples, the commission error is 20 percent. The noise samples constitute a very heterogeneous group of objects that do not adhere to a normal distribution. The omission error is 7 percent, 8 percent, 11 percent, and 14 percent, for Eiker, Kristiansund, Sollihøgda, and Østerdalen, respectively. The corresponding commission errors, including noise samples, are 9 percent, 11 percent, 22 percent, and 33 percent.

TABLE 2. CLASSIFICATION CONFUSION MATRIX

Given Label / True Label	Bright Vehicle	Dark Vehicle	Vehicle Shadow	Road Mark	SUM
Bright vehicle	127	0	0	12	139
Dark vehicle	0	67	10	0	77
Vehicle shadow	0	10	70	0	80
Road marking	0	0	0	3	3
Noise segments	17	16	18	22	73
SUM	144	93	98	37	372

The main objective of our study is to develop methodology for reliable estimation of the amount of traffic on certain roads, thus an important part of the validation is to compare the vehicle counts from the satellite images against manual counts. The manual counts were performed by a person without knowledge about the results of the

automatic counts. Each vehicle was marked in the image. The validation approach consists of comparing the number of vehicles found by manual inspection to the number of vehicles found by the automatic algorithm. The degree of consensus is also checked, i.e., the number of cars that was found by both methods.

In-road equipment counts how many vehicles pass a certain location during some period, which in our case is one hour. In order to compare this number with the counts made by our method, we must apply it to an image that covers the road in an area surrounding the counting station. We select a subset of the road so that the distance on each side of the counting station is maximized, while at the same time, no large road intersections are included. This is necessary since we are going to compare the number of cars in a snapshot in time to the number of cars during a one hour period. The data which is available from the in-road stations include the average vehicle speed. Thus, assuming there is free flow of traffic (no queues) and traffic density is uniformly distributed in the one hour interval, we may predict the number of vehicles that could be expected to appear on the given road segment in the image. This number is not to be regarded as the true number of vehicles in the image. However, the predicted number of vehicles from the in-road sensor gives an idea about how well an image (acquired in a snapshot in time) represents the average traffic situation.

Six different in-road counting stations were used for validation (two in Kristiansund, two in Sollihøgda, one in Eiker, and one in Østerdalen).

The validation results are summarized in Table 3; the number of vehicles counted by the automatic algorithm, the number of vehicles counted by manual inspection, the number of vehicles that exists in both the automatic and the manual counts, and finally, the number of vehicles that is expected to be present in the image based on in-road counts, merely as an interesting comparison. Fragments (class 4) that belong together are counted as one vehicle. Similarly, class 3 (bright truck) has been trained to identify separate trailer wagons. In cases where the truck is pulling two trailers, only one vehicle is counted.

The estimated number of vehicles from in-road equipment does not necessarily represent the truth about the images, but comparing this number to our results is nevertheless interesting, as it gives an idea of how well it is possible to monitor traffic using satellite observations. The manual count is probably the number that is closest to the true number of cars in the image. However, as mentioned above, there are some ambiguities even when it comes to manual counts.

Discussion

Our results indicate that the automatic vehicle detection algorithm tends to underestimate the number of vehicles. In those images where the automatic count is especially low with respect to the manual count (Kristiansund #1 and 2, Eiker, and Sollihøgda #2), the degree of consensus is high. By this we mean that the manual interpreter agreed that most of the vehicles found by the algorithm were in fact vehicles. On the other hand, in those images where the automatic count is close to the manual count (Østerdalen and Sollihøgda #1), the degree of consensus is less, i.e., a relatively large proportion of the segments that were labeled as vehicles by the algorithm was not counted as vehicles by the manual interpreter (Table 3). In these two images, there are a fair number of tree shadows and road marks that are incorrectly classified as vehicles.

The performance of both the segmentation and the classification routine was especially good in the Eiker image (2 percent noise segments). The elevation of the sun at the time of image acquisition was higher in Eiker than for the other locations (Table 1). This result may indicate that such lighting conditions are advantageous, which seems reasonable, considering, e.g., the fact that higher sun elevation means shorter shadows. The lowest omission and commission errors for the classification were obtained for the Kristiansund and Eiker images. These are also the images with highest traffic density. The lowest performance was registered for the Østerdalen image, where the traffic density is very low, and the imaged area consists mainly of woods located next to the road. The density of the traffic is presumably an important factor that affects the performance of the vehicle detection procedure.

As noted for the classification results, most vehicles are classified correctly, thus the main reason for underestimation is believed to lie in the segmentation routine. Some vehicles have very poor contrast and are hard to detect by gray tone level thresholding. We also see examples of vehicles that partly or fully lie in shadow (Figure 10). Furthermore, some vehicles are lost since segments that overlap the road edge are discarded. However, the opposite strategy (not excluding road edge segments) leads to drastic increase in false positives, as there is a large amount of tree shadows along the type of roads that we are interested in. In future work, alternative ways of removing tree shadows should be considered. Tree shadows could possibly constitute a separate class, although their size and shape varies considerably.

Segments that do not belong to any of the six defined classes also pose a great challenge to the automatic algorithm. These segments include road marks of various shapes and

TABLE 3. VALIDATION RESULTS. THE COLUMN "AGREEMENT" INDICATE THE NUMBER OF VEHICLES THAT WERE FOUND BY BOTH THE AUTOMATIC ALGORITHM AND THE PERSON WHO COUNTED VEHICLES MANUALLY. THE RIGHTMOST COLUMN INDICATES HOW MANY VEHICLES ARE EXPECTED TO BE PRESENT IN THE IMAGE IN A SNAPSHOT OF TIME GIVEN THE NUMBER OF VEHICLES THAT PASSED THE IN-ROAD STATION DURING A ONE HOUR PERIOD AND THE LENGTH OF THE ROAD SEGMENT IN THE IMAGE

			Number of Vehicles			
Location	Length of Road Segment (m)	Time of Image Acquisition (UTC)	Automatic Count	Manual Count	Agreement	Estimated from In-road Count 10–11 UTC
Kristiansund # 1	1055	10:56	16	22	14	25
Kristiansund # 2	5775	10:56	18	32	17	27
Østerdalen	31779	10:39	43	44	30	51
Eiker	7836	10:42	39	57	36	57
Sollihøgda # 1	7819	10:32	63	64	48	58
Sollihøgda # 2	6139	10:32	26	30	24	38

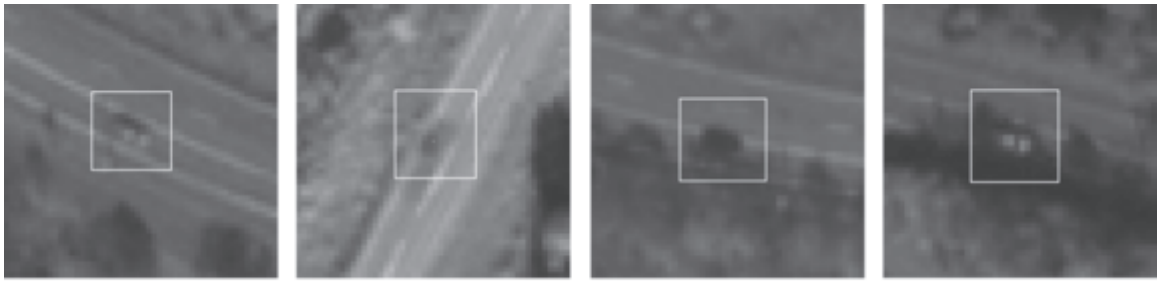

Figure 10. The images show examples of vehicles that were counted manually, but not detected by the automatic algorithm. (a) and (b): Vehicles that have low contrast to the local background, and (c) and (d): Vehicles partly hidden behind trees/shadows.

with high contrast, tree shadows, parts of road signs, bridges, roundabouts, and so on; i.e., they constitute a highly heterogeneous group of segments that demand special treatment when it comes to classification (Figure 8).

The segmentation step was developed using image data from three locations with relatively good lighting conditions. The hypothesis about the relationship between the image histogram and the intensity classes was therefore too optimistic. Studies of a larger set of images reveal that variations in lighting conditions sometimes yield less desirable results. Further research should be made in order to adapt the segmentation algorithm to different lighting conditions. An alternative that should be considered in the future is to apply some sort of preprocessing filter to the image in order to enhance the contrast of the vehicles before segmentation. In Alba-Flores (2005) it is suggested to apply a filter that assigns to each pixel the maximum or minimum intensity value in a neighborhood of 3-by-3 pixels, before using Otsu's threshold selection for bright or dark objects, respectively.

Conclusions

The monitoring of traffic conditions is necessary for development and maintenance of the road network. Deriving traffic statistics involves estimating the Annual Average Day Traffic (AADT), which is a measure of the traffic distribution during a day at a specific point and averaged over a year. The current method for obtaining data for AADT estimation is based on a limited number of in-road counting stations with expensive equipment. Counting vehicles from space, using satellite imagery, may provide supplementary information about the traffic situation and improved spatial coverage. We have presented an automated approach for vehicle detection in very-high-resolution satellite images. The method consists of a segmentation step followed by object based feature extraction and classification. Improvements are needed before the methodology can be used operationally, thus the presented results are preliminary, and further work is ongoing. According to the Norwegian Public Roads Administration the results are promising for the improved monitoring of the national road system in the future.

Past research on automatic vehicle detection from images has focused significantly on the use of aerial imagery, although a few studies using satellite images have been reported. In our study, we have aimed at making as few assumptions as possible about the image data for which the vehicle detection algorithm is to be applied. Our contribution includes a proposed method for how to construct a vehicle shadow mask, i.e., a mask containing dark segments that are likely to represent a shadow. Vehicle shadows provide valuable information about the segments, and should be exploited for detection purposes. In our case, we have integrated the shadow information during the classification step.

The proposed method finds image segments representing potential vehicles by looking for areas that are darker or brighter than their surroundings. These segments are then described using a set of features. The features are values that represent various properties regarding the geometric shape and the intensity contrast of the object. According to their feature values, the classification gives each segment a label that represents one of the six defined classes (Bright car, Dark car, Bright truck, Bright vehicle fragment, Vehicle shadow, or Road mark (arrow)).

Approximately 20 percent of the total number of segments resulting from segmentation do not belong to any class, and out of these segments, 28 percent are erroneously classified as vehicles. However, in general, the suggested approach underestimates the number of vehicles in the image. Vehicles are usually missed due to local variations in lighting conditions or obstructed view of the road due to vegetation. The main source of non-vehicle segments that are confused with vehicles is road segments, which (after segmentation) come in a wide variation of shapes and shades.

It is desirable to be able to discriminate a larger part of the noise objects, i.e., objects that do not belong to any of the defined classes. Perhaps we need to use specific features that are able to separate outlier segments from the classes, e.g., context-based features. Furthermore, a more thorough pre-classification may exclude outliers on an early stage and ease the task of statistical classification.

The proposed segmentation routine fails to capture vehicles of very low contrast to the local background, especially when the low contrast segments are only slightly brighter than the road. Lowering the threshold for bright segments may yield a lot more road marks to become included as segments. Alternative methods that should be tested include sophisticated region growing techniques, where pixels of low contrast are expanded into regions only if the resulting object meets certain requirements.

Acknowledgments

This work was funded by the Norwegian Public Roads Administration, the Norwegian Space Centre, and the Research Council of Norway. We would also like to thank

our colleagues; Jostein Amlien for manual vehicle counting and Line Eikvil for helpful comments.

References

Aurdal, L, L. Eikvil, H. Koren, J.U. Hanssen, K. Johansen, and M. Holden, 2007. *Road Traffic Snapshot*, Report 1015, Norwegian Computing Center, Oslo, Norway.

Alba-Flores, R., 2005. *Evaluation of the Use of High-resolution Satellite Imagery in Transportation Applications, Final Report*, CTS 05–11, Department of Electrical and Computer Engineering, University of Minnesota, Duluth, Minnesota.

Burlina, P., V. Paramerswaran, and R. Chellapa, 1997. Sensitivity analysis and learning strategies for context-based vehicle detection algorithms, *Proceedings of DARPA Image Understanding Workshop*, May, New Orleans, Louisiana, pp. 577–584.

Eikvil, L., L. Aurdal, and H. Koren, 2008. Classification-based vehicle detection in high-resolution satellite images, *ISPRS Journal of Photogrammetry and Remote Sensing*, In press, doi:10.1016/j.isprsjprs.2008.09.005.

Gerhardinger, A., D. Ehrlich, and M. Pesaresi, 2005. Vehicles detection from very high resolution satellite imagery, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, Part 3/W24, pp. 83–88.

Gonzalez, R.C., and R.E. Woods, 1992. *Digital Image Processing*, Addison-Wesley Publishing Company, 716 p.

Hinz, S., 2005. Detection of vehicles and vehicle queues for road monitoring using high resolution aerial images, *Proceedings of 9th World Multiconference on Systemics, Cybernetics and Informatics*, 10–13 July, Orlando, Florida, unpaginated CD-ROM.

Hinz, S., J. Leitloff, and U. Stilla, 2005. Context-supported vehicle detection in optical satellite images of urban areas, *Proceedings of IGARSS '05, IEEE International Geoscience and Remote Sensing Symposium*, 25–29 July, Seoul, Korea, Vol. 4, pp. 2937–2241.

Jin, X., and C.H. Davis, 2004. Vector-guided vehicle detection from high-resolution satellite imagery, *Proceedings of IGARSS '04, IEEE International Geoscience and Remote Sensing Symposium*, 20–24 September, Anchorage, Alaska, Vol. 2, pp. 1095–1098.

Jin, X., and C.H. Davis, 2007. Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks, *Image and Vision Computing*, 25(9):1422–1431.

Maitra, S., 1979. *Moments invariants*, *Proceedings of the IEEE*, 67(4):697–699.

Otsu, N., 1979. A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66.

Pesaresi, M., K.H. Gutjahr, and E. Pagot, 2008. Estimating the velocity and direction of moving targets using a single optical VHR satellite sensor image, *International Journal of Remote Sensing*, 29(4):1221–1228.

Schlosser, C., J. Reitberger, and S. Hinz, 2003. Automatic car detection in high resolution urban scenes based on an adaptive 3D-model, *Proceedings of the IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, Berlin, Germany, pp. 167–171.

Sharma, G., C.J. Merry, P. Goel, and M. McCord, 2006. Vehicle detection in 1-m resolution satellite and airborne imagery, *International Journal of Remote Sensing*, 27(4):779–797.

Zhao, T., and R. Nevatia, 2001. Car detection in low resolution aerial images, *Proceedings of the 8th IEEE International Conference on Computer Vision*, 09–10 July, Vancouver, Canada, pp. 710–717.

Zheng, H., L. Pan, and L. Li, 2006. A morphological neural network approach for vehicle detection from high resolution satellite imagery, *Proceedings of the International Conference on Neural Information Processing, Lecture Notes in Computer Science* (I. King, J. Wang, L. Chan, and D.L. Wang, editors), Vol. 4233, Springer, pp. 99–106.

Zheng, H., and L. Li, 2007. An artificial immune approach for vehicle detection from high resolution space imagery, *International Journal of Computer Science and Network Security*, 7(2):67–72.

(Received 28 July 2008; accepted 05 November 2008; revised 20 January 2009)