MAPPING ROAD TRAFFIC CONDITIONS USING HIGH RESOLUTION SATELLITE IMAGES

S. Ø. Larsen [a, *], J. Amlien[a], H. Koren[a], R. Solberg[a]

[a] Norwegian Computing Center, P.O.Box 114 Blindern, 0314 Oslo, Norway
- siri.oyen.larsen@nr.no, +47 2285 2615

KEY WORDS: Remote Sensing, Vehicle Detection, Pattern Recognition, Traffic Statistics, High-Resolution Satellite Images, Object-Based Segmentation, QuickBird.

ABSTRACT:

Construction, development and maintenance of the road network are central activities for several public authorities. In cooperation with the Norwegian road authorities, we have developed an approach for automated vehicle detection and generation of traffic statistics from QuickBird images. Satellite surveillance serves several obvious advantages over the methods that are being used today, which consist of expensive single-point measurements made from pressure sensors, video surveillance etc., in or close to the road. Based on advice from the road authorities of Norway, we have selected a set of study sites from different parts of the country, such that our image data represents the diversity of road types and solar illumination conditions. Road and vegetation masks are applied to the image so that the search for vehicles is restricted to the (paved parts of the) roads only. For segmentation, we have applied techniques that seek to locate the modes of the image histogram. The resulting segments are then examined by feature extraction and classified adopting the maximum likelihood method. Additionally, we propose a new approach for car shadow removal. The described methods are implemented and tested against manual vehicle counts. We also compare the results to area traffic cover statistics estimated from single-point measurements. Manual vehicle counts indicate that there is some ambiguity in the interpretation of the images. Nevertheless, the automatic method that we have developed in this study performs very well compared with the reported manual counts.

1. INTRODUCTION

1.1 Background and objectives

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 Day Traffic (ADT), 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. A potential alternative is traffic counting in very-high-resolution satellite images, like QuickBird (0.6 m resolution). 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 automatic image analysis methodology – pattern recognition tailored to the detection of vehicles.

The main objective of our study is to develop methodology for automatic vehicle detection using very-high-resolution satellite imagery, i.e., spatial resolutions of 0.6-1.0 m in the panchromatic band, which is available from the new generation of high-resolution commercial satellite images. In cooperation with the local road authorities, the aim was to compare ADT measurements with the traffic statistics that were estimated from automatic counts in QuickBird images.

1.2 Related work

Extensive research has been performed on vehicle detection in aerial imagery, (e.g., Burlina, 1997; Hinz, 2005; Schlosser, 2003; Zhao, 2001). However, very few studies address vehicle detection from very-high-resolution satellite images. Jin and Davis detect vehicles in Ikonos (1.0 m) images (Jin, 2007). Their approach is based on an implicit vehicle model, and the classification is performed on the pixel level. The training data is derived from manual delineation of vehicles in the images. Sharma presents three different approaches for vehicle detection in Ikonos images (Sharma, 2006). Of these, the best performance is obtained using a pixel-based Bayesian Background Transformation approach, which requires the existence of a high quality background estimate. Alba-Flores detects vehicles in Ikonos images of US one-way highways using two different thresholding approaches, (Alba-Flores, 2007). The work presented in our paper is based on the ESA (European Space Agency) project “Road Traffic Snapshot”, which concerns a possible future service for counting vehicles in satellite images and generating traffic information based on these counts (Aurdal, 2007).

2. EXPERIMENTAL DATA

To be able to detect vehicles, satellite images with high resolution are required. We have chosen the QuickBird satellite with 0.6 m resolution panchromatic band. Our data consists of

* Corresponding author.

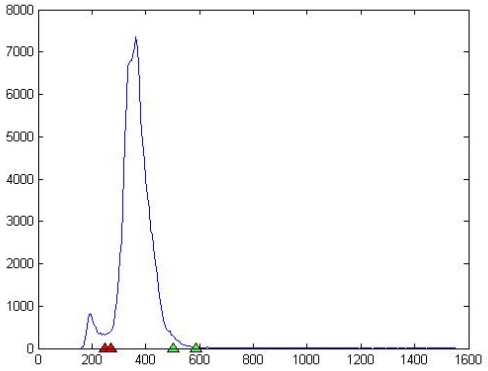

Figure 1. Typical image histogram for (a subset of) a panchromatic image after application of road and vegetation masks.

five subsets of satellite images from the QuickBird image archive, and covers different parts of Norway for the period between 2002 and 2006. 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 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. Traffic counts from in-road equipment measured at the same dates and times for the corresponding roads were provided by the Norwegian Public Roads Administration, and were used for comparison with the image analysis based counts.

3. METHODOLOGY

3.1 Road and vegetation masks

The problem of finding vehicles in an image is substantially reduced in the event that the position of the road is known. We restrict the search for vehicles to the parts of the image representing paved road.

Methods for both manual and automatic generation of road masks were investigated. See (Larsen, 2008) for a thorough description of this topic. Automatic generation of road masks faces certain difficulties; tunnels and crossing bridges should be removed, all gaps have to be closed, roundabouts have to be included, etc. In the rest of this paper we will present methods that have been developed on images after application of *manually* delineated road masks.

In some cases, vegetation covers parts of the roads, typically tree canopy hiding parts of the road, or vegetation growing between the lanes of the road. Consequently, we also apply a vegetation mask. The vegetation mask is retrieved from multispectral imagery, which is provided by the QuickBird sensor at a resolution of 2.4 m, and where each resolution cell is referenced to 4×4 cells in the corresponding panchromatic image. More specifically, the normalized difference vegetation index (NDVI) is computed from the red and infrared image bands, after resampling to the resolution of the panchromatic image, using cubic interpolation. The vegetation mask is produced by thresholding the NDVI image. The appropriate threshold is found using Otsu's algorithm (Otsu, 1979).

3.2 Segmentation

The main segmentation routine is based on finding segments that are darker or brighter than their surroundings and is applied to the masked panchromatic image. In the image histogram (Figure 1), the road (asphalt) pixels constitute the dominating histogram mode. Note that the background pixels are left out of 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.

The panchromatic image is thresholded in two stages, for dark and bright segments, resepectively. We apply Otsu's method for unsupervised threshold selection (Otsu, 1979). Note that Otsu's method easily may 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 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).

For each stage of the segmentation, i.e., the dark and the bright stage, *two* thresholds were found. One threshold is strict, the other threshold is loose. Each threshold is found by restricting the focus to a subset of the image histogram. First, we locate the main peak of the histogram, simply by finding the mean value.

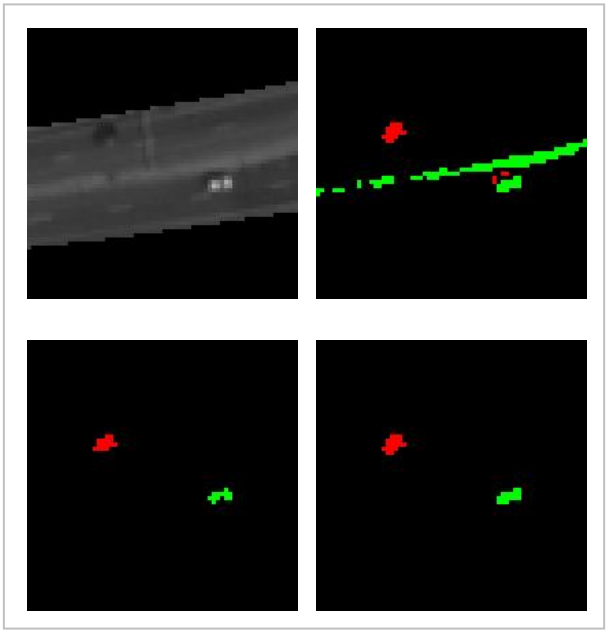

Figure 2. Segmentation stages. Upper left: masked panchromatic image;. upper right: the result of segmentation with the loose threshold for bright (green) and dark (red) objects;. lower left: corresponding result using the strict thresholds; lower right: the combined result.

The image is thresholded in two steps, using the strict and the loose thresholds separately. The two results are then combined by a kind of hysteresis thresholding; a segment resulting from the loose threshold is only kept if it contains a segment resulting from the strict threshold. Finally, the results from the dark and bright segmentation stages are combined.

More specifically, 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 define the thresholds as follows:

- Segmentation of dark segments:
  - strict threshold: Otsu applied to the histogram on the interval $[I_{min}, \mu - \sigma]$
  - loose threshold: Otsu applied to the histogram on the interval $[I_{min}, \mu - 0.5\sigma]$
- Segmentation of bright segments:
  - loose threshold: Otsu applied to the histogram on the interval $[\mu + \sigma, I_{max}]$
  - strict threshold: $\mu + 3\sigma$

Segmentation using a loose threshold tends to produce too many segments, while the strict threshold often yields poorly defined or fragmented segments. The combination of the two thresholds helps to provide more well-defined objects, while at the same time ignoring objects that are only slightly different from the background asphalt color, and therefore most likely not a vehicle (Figure 2). For illustration purposes, only a small part of the image is shown in the figure.

3.3 Shadows

Segmentation based on intensity values alone inevitably produces some unwelcome segments. Vehicle shadows, as well as cast shadows by trees along the road, are the main contributors to dark non-vehicle segments that need to be removed.

Roadside shadows

Cast shadows from trees or other objects at the roadside are removed using the simple assumption that no real vehicle segments should be located on the outer edge of the road. We compute a road-edge mask from the manually drawn road mask, using dilation of the road mask with a suitable structuring element. The resulting edge mask is very narrow. Any segment overlapping the road-edge mask is regarded a non-vehicle segment, and discarded.

Vehicle shadows

A vehicle shadow, especially when appearing on the longer side of a car, is visually very similar to a dark vehicle. It is usually easy to separate a dark vehicle from a vehicle shadow by manual inspection, as the shadow must obviously reside next to the vehicle that casts it. In the QuickBird images that we have studied, vehicle shadows are only visible next to bright vehicles. Visually it is impossible to distinguish a dark vehicle from its shadow.

We use information about the lighting and viewing conditions that existed during image acquisition in order to detect dark segments that are positioned in the *expected* shadow zone of bright segments. The direction and length of the shadow sector are estimated from the sun azimuth and sun elevation, respectively, at the time of image acquisition.

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. The length of the shadow (half the size of the structure element), is given by the average vehicle height divided by the tangent of the sun elevation angle. For QuickBird images we use an average vehicle height of 3 pixels, corresponding to 1.8 meters.

We then create a structure element which represents the expected shadow zone that applies to the given image, and use it to perform dilation of the bright segment image. The dilation result represents an image of the bright objects together with their expected shadow zones (Figure 3b). The bright object segment image is subtracted from the result of dilation, yielding an image representing the expected shadow zones only. This image is then compared to a segmented image of dark objects situated close to a bright object. The latter image is found by first calculating the distance map of the bright segment image. This distance map is thresholded at a low value, and the result is multiplied with the dark segment image. Finally, we compare the expected shadow zone image with the image of dark segments that are located close to bright segments. Whereever there is overlap between these two images, the dark object is assumed to be a shadow (Figure 3).

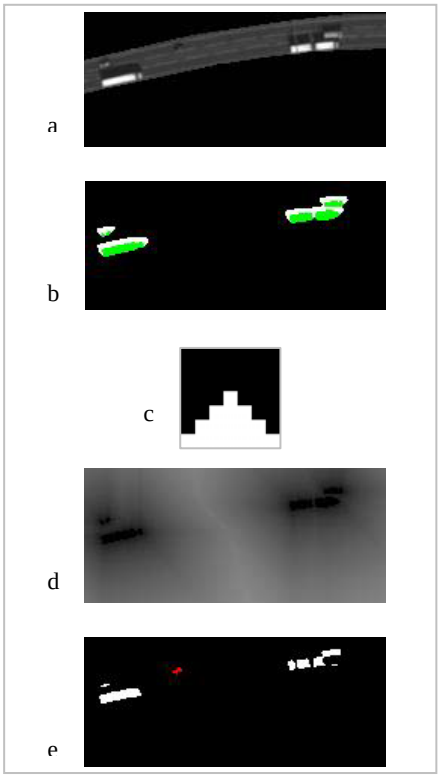

Figure 3. a) Panchromatic image, vehicles and shadows. b) Bright segments (green) and expected shadow zone (white). c) The structure element used for dilation. d) Distance map – bright segments. e) Dark segments – assumed to be shadows (white) and assumed to be vehicles (red).

3.4 Classification

As already mentioned the segments produced by the segmentation routine must be treated further because they include non-vehicle segments. We use a maximum likelihood approach and classify the segments into six classes (four vehicle classes + two non-vehicle classes):

1. Bright car
2. Dark car
3. Bright truck
4. Brigh vehicle fragment
5. Vehicle shadow
6. Road marking

According to Bayesian decison theory, the probability of error is minimized if we assign an object with a given feature vector to the class for which the posterior probability of the vector belonging to this class is maximized (over all classes). This method is (among other names) called maximum likelihood, or minimum-error-rate classification (Duda, 2000). The maximum likelihood classification states that a feature vector ξ should be classified to class κ if the posterior probability $P(\omega_\kappa | \xi)$ – the probability that the correct class is class κ, given that the feature vector ξ has been measured – is maximized, i.e., $P(\omega_\kappa | \xi) > P(\omega_\beta | \xi)$ for all $\beta \neq \kappa$. The posterior probability is given by Bayes formula $P(\omega_\kappa | \xi) = p(\xi | \omega_\kappa)P(\omega_\kappa) / p(\xi)$, where $p(\xi | \omega_\kappa)$ is the conditional probability of measuring feature vector ξ, given that the vector is a sample from class κ, $P(\omega_\kappa)$ is the prior probability of class κ, and $p(\xi)$ is the evidence $p(\xi) = \sum_{\beta=1}^{K} p(\xi | \omega_\beta) P(\omega_\beta)$, where K is the number of classes.

It is natural to assume that the feature vectors belonging to a given class have a Gaussian distribution, i.e., the conditional probability $p(\xi | \omega_\kappa)$ is a (multivariate) normal distribution. Studies of scatterplots indicate that it is not reasonable to assume that the classes have equal covariance matrices. Furthermore, the features are assumed to be correlated, i.e., have general (non-diagonal) covariance matrices.

The prior probability of a given class is the probability that a random object, having no other information about its shape etc., belongs to this class. Common choices include equal, constant probabilities for the classes or the use of class frequencies. We use the latter, i.e., 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.

Feature extraction

Various features may be used in order to describe the shape and spectral characteristics of segments. A number of features were examined, using feature selection methods together with the available training data. The best performance of the maximum likelihood classification was obtained with the following set of six features:

- the mean of the region intensity
- the mean of the region gradient (computed by the Sobel operator)
- the standard deviation of the region intensity
- the length of the region's bounding box
- the 1st Hu moment of the region
- the spatial spread of the region

In addition to the above features, the following features are used for pre- or post-classification:

- region area
- region elongation
- distance to nearest vehicle shadow (assuming a vehicle shadow mask as given by the technique described in Section 3.3)

Preclassification

Before the statistical classification, we perform an initial, rule-based classification in order to discard segments that are obvious non-vehicle segments. The region area must neither be too large nor too small. We also require that the elongation adhere to a given interval. Furthermore, we assume that the mean intensity value within the region is above or below preset thresholds, and the region gradient must also obey a minimum threshold value.

Postclassification

Vehicle shadows are often confused with dark vehicles and vice versa, even after application of the maximum likelihood classificator. 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 postclassification is based on the "distance to nearest shadow" feature, calculated from the vehicle shadow mask, as described in Section 3.3. Specifically, the class label of a dark vehicle is changed to shadow if its shadow distance is zero.

The distance-to-shadow information is also used to improve the classification of bright vehicle fragments into road markings. 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 bright vehicle fragment if its distance to a shadow is less than three pixel units.

4. CLASSIFICATION RESULTS

The classification method was trained and tested on a selected set of subimages. Each of the images in the training and test sets were processed through the following steps: application of road and vegetation mask, segmentation (including shadow masks), feature extraction, and preclassification, before they were manually labeled. We used labels 1-6 for the respective classes (see section 3.4), and a separate label for segments that should be rejected. Only segments belonging to one of the six classes were included in the training set. 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. For testing, we had 310 samples. The results are given in Table 1. Not including the reject segments, the classification rate is 88.7%. Furthermore, 89.6% percent of the vehicles were classified as vehicles, while 70.1% percent of non-vehicles were classified as non-vehicles.

There is a considerable amount of segments that do not belong to any of the six classes (Table 1). By inspecting the images and the segmentation results, we find that these segments constitute a very heterogeneous group of objects. The maximum likelihood classification approach might be applied with a reject

Given label / True label	Bright vehicle	Dark vehicle	Vehicle shadow	Road mark	SUM
Bright vehicle	96	0	0	11	**107**
Dark vehicle	0	59	7	0	**66**
Vehicle shadow	0	10	62	0	**72**
Road marking	0	0	0	2	**2**
Reject	11	20	22	10	**63**
SUM	**107**	**89**	**91**	**23**	**310**

Table 1. Classification results.

option; if the posterior probability does not exceed a given level for neither of the classes, the segment should be rejected. This approach was tested, but did not yield any improvements.

5. VALIDATION

One of the main objectives of our study was to develop methodology that is able to make reliable estimates of the amount of traffic on certain roads. One important part of the validation has therefore been to compare the vehicle counts from the satellite images against counts made by the in-road equipment. Secondly, our results have been manually assessed.

In-road equipment counts how many vehicles that pass a certain location during 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, we may estimate the number of vehicles that should appear on the given road stretch in the snapshot in time when the satellite image was taken. Furthermore, we perform the opposite comparison; estimate the expected number of cars to pass a single location in space during a period of one hour given the number of vehicles that was counted in the satellite image.

Even with a resolution of 0.6 m, counting vehicles in a satellite image by manual inspection is a task that may yield ambiguous results. Thus, the validation should be performed by more than one individual. In our study, two different persons have provided manual counts. These counts may be compared directly to the counts that were generated by the automatic routine. Furthermore, the degree of consensus must be checked.

Six different counting stations were used for validation. Table 2 summarizes the results. Some comments should be made regarding the numbers in the table. First of all, note that since we have a class called “bright vehicle fragments”, the method will sometimes locate two fragments of the same vehicle. However, the two fragments should be counted as one vehicle. Similarly, the class “bright truck” has been trained to identify one trailer wagon. In cases where the truck is pulling two wagons, only one vehicle should be counted. In Table 2 the column “Objects classified as vehicles” presents the number of vehicles found by the algorithm counting fragments that belong together only once. The numbers in parenthesis correspond to the number of segments that were classified into one of the vehicles classes, i.e., bright car, dark car, bright truck, or bright vehicle fragment.

In the Eiker image a large stretch of road lies in the shadow of a large cloud. The automatic method did not find any vehicles on this stretch. Both person 1 and person 2 located ten vehicles on this stretch.

The most important sources of non-vehicles classified as vehicles are road markings and tree shadows (Figure 4). In the Østerdalen image, there are 29 tree shadows and eight road markings that are wrongly classified as vehicles. In the Sollihøgda # 1 image, the corresponding numbers of tree shadows and road markings are three and twelve. As described above, a road edge shadow mask was applied in order to remove as many tree shadows as possible before classification.

6. DISCUSSION AND CONCLUSIONS

We have presented an approach for segmentation and classification of vehicles from high-resolution satellite images. The method has been developed using a selected set of images of Norwegian roads, but we have attempted to make it as general as possible. The method is object based, and applies well-known pattern classification techniques to segments that are found in the Otsu-inspired segmentation step.

Many dark segments on the road are shadows of bright vehicles. We propose a method for how to construct a vehicle shadow mask, i.e. a mask containing those dark segments that are likely to represent a shadow. The mask is helpful during the classification step, as the shadows provide valuable information about the segments. For instance, the classifier often confuses the classes “bright vehicle fragment” and “road marking”. Since we know that road markings do not cast shadows, while bright vehicle fragments often do, the “distance to nearest shadow” feature may be used to redirect the classification result. Of course, this approach depends on the confidence of the vehicle shadow mask. If too many dark vehicles are included in the shadow mask, the approach will fail. In our test data, 62 out of 72 vehicle shadows, and 59 out of 66 dark vehicles, were correctly classified – a result that we consider as satisfactory.

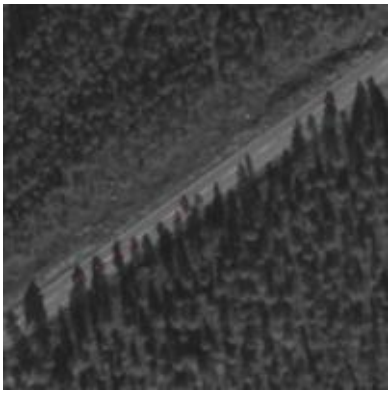

Figure 4. Tree shadows stretching far into the road.

Location	Time of image acquisition UTC	Manual vehicle count			Objects classified as vehicles	Correctly classified as vehicles		Estimated vehicle count from in-road station 10-11 UTC
		Person 1	Person 2	Consensus		Person 1	Person 2	
Kristiansund # 1	10:56	22	22	21	17 (17)	14	14	25
Kristiansund # 2	10:56	32	33	32	22 (25)	21	21	27
Østerdalen	10:39	44	41	41	80 (88)	32	32	51
Eiker	10:42	57	55	54	39 (44)	35	32	57
Sollihøgda # 1	10:32	63	63	62	64 (65)	48	47	58
Sollihøgda # 2	10:32	30	30	30	26 (27)	24	24	38

Table 2. Manual vs. automatic vs. in-road counts.

As noted in section 5, objects at the roadside, especially trees, introduce a great challenge to the automatic vehicle counting method. A special example is the Østerdalen image, where 29 tree shadows are counted as vehicles. Subtracting 29 from the automatic count, the result is very close to the manual counts as well as the number of vehicles estimated by in-road equipment counts. The mentioned image is an example of where the tree crowns enter quite far into the road. The shadows of the trees are therefore not located at the edge of the road, but rather on the edge of the vegetation mask. It may be discussed whether the road edge mask should be constructed from the total mask, i.e., the road mask combined with the vegetation mask, instead of the road mask only.

Although various features have been tested, the number could be extended. It is desirable to be able to discriminate a larger part of the reject objects from the rest. Perhaps we need to use specific features that are able to separate outlier segments from the classes, e.g., context-based features.

The performance of our algorithm may be summed up as follows: The majority of vehicles that are found in the segmentation step are correctly classified as vehicles. Tree shadows that enter far into the road, as well as high contrast road marks tend to be confused as vehicles. The segmentation routine should be improved so that it finds even vehicles with low contrast. Alternative features and context-based information may help to reduce the number of false vehicle counts.

ACKNOWLEDGEMENTS

We thank our colleagues Øivind Due Trier and Ragnar Bang Huseby for helpful comments. The software code for algorithm testing was written as a further development of the RoadTraffic software tool (Aurdal, 2007). The work presented in this paper was funded by the Norwegian Public Roads Administration and the Norwegian Space Centre.

REFERENCES

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, USA.

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.

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*, pp. 577-584.

Duda, R. O., P. E. Hart, and D. G. Stork, 2000. *Pattern Classification*. Wiley-Interscience, pp. 20-44.

Eikvil, L., L. Aurdal, and H. Koren, 200?. Classification based vehicle detection in high-resolution satellite images. Submitted to *ISPRS Journal of Photogrammetry and Remote Sensing.*

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*, Orlando, Florida, USA.

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), pp. 1422-1431.

Niu, X., 2006. A semi-automatic framework for highway extraction and vehicle detection based on a geometric deformable model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(3-4), pp. 170-186.

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

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.

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), pp. 779-797.

Zhao, T. and R. Nevatia, 2001. Car detection in low resolution aerial images. *Proceedings of the International Conference on Computer Vision*, Vancouver, Canada, pp. 710-717.