ISPRS Journal of Photogrammetry and Remote Sensing 64 (2009) 65–72

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Classification-based vehicle detection in high-resolution satellite images

Line Eikvil *, Lars Aurdal, Hans Koren

Norwegian Computing Center, P.O. Box 114, Blindern, NO-0314 Oslo, Norway

A R T I C L E I N F O

Article history:
Received 19 December 2006
Received in revised form
19 August 2008
Accepted 3 September 2008
Available online 30 October 2008

Keywords:
Quickbird
Vehicle detection
Classification

A B S T R A C T

In this study, we have looked into the problem of vehicle detection in high-resolution satellite images. Based on the input from the local road authorities, we have focused not only on highways, but also on inner city roads, where more clutter is expected. The study site is the city of Oslo, Norway. To do vehicle detection in these areas, we propose an automatic approach, consisting of a segmentation step, followed by two stages of object classification. In the process, we utilize multispectral images, panchromatic images and a road network. The approach has been tested on Quickbird images, and the results that are obtained have been compared with manual counts and classifications.

1. Introduction

Information on vehicle location and movement is important for transport planning, accessibility analyses and traffic statistics. Monitoring of traffic is therefore an important task for the authorities. Current technology typically collects data by the use of *inductive loop vehicle detectors* embedded in (or lying on) roadways. This equipment provides traffic flow information over time for a point in space. However, other types of information from a larger area could often be useful to better understand the dynamics of the traffic. Images of a large road network could, for instance, be used to acquire information from a whole region at one point in time. Such snapshots of entire networks could give more insight into the distribution of vehicles in a region, and could also provide valuable information for areas not covered by traditional counting equipment. Until recently acquisition of such traffic snapshots has not been feasible, but since the launch of civil optical high-resolution satellite systems like Ikonos and QuickBird, images with a resolution of around one meter are now available. Hence, this type of imagery has a resolution that may make it possible to extract road traffic information.

In this study, we have looked into the problem of vehicle detection in high-resolution satellite images. Based on input from local road authorities, we will not focus only on highways, as has been done in a few other studies (Alba-Flores, 2005; McCord et al., 1998; Sharma, 2002; Sharma et al., 2006), but also on inner city roads, where more clutter is expected. The study site is the city of Oslo, Norway, where three different areas have been selected. Here, the roads will be narrower than typical highways, and more vegetation along the roads is expected. In addition, the high latitude of the study area affects the light conditions. To do vehicle detection in these areas, we propose an automatic approach consisting of a segmentation step followed by object classification, utilizing multispectral images, panchromatic images and road network information.

2. Background

The main methodological challenge is related to the problem of vehicle detection in high-resolution satellite images, where the resolution on the ground is low compared with the size of the objects sought for analysis. Compared with vehicle detection in aerial images, this poses special problems and will, for instance, make it more difficult to separate vehicles from other types of object such as trees, road markings etc.

For vehicle detection in aerial images, different approaches have been investigated (Hinz, 2004; Schlosser et al., 2003; Stilla et al., 2004; Toth and Grejner-Brzezinska, 2005; Zhao and Nevatia, 2001). In these images, the resolution is higher than in satellite images and typical vehicles have a length of 13–26 pixels (Zhao and Nevatia, 2001). Hence, more details of the vehicles are visible in these images, and many of the approaches used for detection have applied 3D models (Hinz, 2004; Schlosser et al., 2003; Stilla et al., 2004; Zhao and Nevatia, 2001). Detection rates are generally high at this resolution. Less work has been done on vehicle detection in satellite images, but a few studies exist and a few methods are suggested for analysis of the panchromatic images.

Sharma et al. (2006) detect vehicles on highways in 1-m resolution IKONOS images. Three different approaches are used: (i) attempts to maximize the separation between vehicles and road surface by the use of principal component analysis (PCA), and thresholding of one PCA band, (ii) uses a Bayesian Background

* Corresponding author. Tel.: +47 22 85 25 00; fax: +47 22 69 76 60.
E-mail addresses: Line.Eikvil@nr.no (L. Eikvil), Lars.Aurdal@nr.no (L. Aurdal), Hans.Koren@nr.no (H. Koren).

0924-2716/$ – see front matter © 2008 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2008.09.005

Transformation (BBT) to classify pixels as stationary or dynamic, by comparing the image to a historic background estimate, while (iii) uses a gradient filter followed by thresholding, morphological operations and clustering to detect vehicle objects. For all approaches, filtering on size, orientation etc. of the resulting connected components is performed. The approaches have all been tested on US highways. Of the different approaches the BBT method is reported to show the best overall performance. This method does however require that a good quality background estimate exists, which can often be a problem to obtain automatically.

Alba-Flores (2005) detects vehicles on highways in IKONOS images using two different thresholding approaches (multiple thresholds and Otsu's method (Otsu, 1979)). The approach is limited to one-way highway segments, and it has been tested on highways in the US. The use of pattern recognition techniques is proposed to make the detection more robust, but has not been applied. Gerhardinger et al. (2005) propose an inductive learning approach (i.e. a learning by example approach — where the system tries to induce a general rule from a set of observed instances). The approach uses characteristics such as radiometry, size, position and pixel patterns to distinguish vehicles from other objects within the road segments and it has been tested for highways in Baghdad.

Leitloff et al. (2005) have looked at the specific problem of detecting queues in urban areas. Regions of interest derived from data fetched from a geographical information system (GIS) are used and vehicles are detected in Quickbird images within these regions of interest. Queues are extracted by detecting lines within these areas. Single vehicles are then detected within the extracted line (queue), by detecting points of minimum and maximum width along the line.

Most of these approaches have been tested on highways in open areas, while we need an approach that will work also for inner-city roads. Here more clutter is expected, and we have therefore focused on finding a classification-based approach to be able to distinguish vehicle objects from different types of clutter. For this reason, we also want to exploit multispectral information, which can help in separating vegetation from other objects.

We will perform vehicle detection and classification in Quickbird images. The panchromatic band of these images has a resolution of 0.6×0.6 m. This means that an average-sized vehicle of 1.7×4 m will cover 2.8 by 6.7 pixels. Multispectral bands have a resolution of 2.44 m. Hence, in the multispectral images an average-sized vehicle will only have a size of 0.7×1.6 pixels. In addition, there is a small time delay between the acquisition of the panchromatic and the multispectral image. This means that for moving vehicles, there will not be a correspondence between the position found in the panchromatic image and that found in the multispectral image. For actual vehicle detection we have therefore chosen to use the panchromatic band only. However, multispectral bands provide information which is useful, for instance, for distinguishing areas of vegetation from the road surface.

3. Methods

In the following we will describe the classification-based approach that has been developed. The approach consists of segmentation, feature extraction, pre-classification and a final classification.

3.1. Segmentation

3.1.1. Segmentation of vegetation

Trees shadowing the road can be a problem for vehicle detection, both because vehicles may be hidden behind the trees and because trees may be confused with vehicles. The first problem cannot be solved, but the second problem can be helped by segmenting trees (and other vegetation) by utilizing multispectral information. We have done this by computing the NDVI (normalized difference vegetation index) from the multispectral data. By thresholding the resulting NDVI image, the areas with vegetation can be identified and masked out.

From a multispectral Quickbird image, we have performed a resampling to the resolution of the panchromatic image using cubic interpolation. From the resulting image, NDVI was computed as: $(\mathrm{NIR}-\mathrm{R})/(\mathrm{NIR}+\mathrm{R})$. Otsu's method for threshold selection (Otsu, 1979) was then applied to the NDVI image to obtain a vegetation mask.

3.1.2. Segmentation of shadowed areas

For some parts of the road network, larger areas may be very dark due to shadows from tall buildings. These areas need to be identified and specially analysed. In this study, we have only focused on identifying these areas, while the problem of analysing the contents has been left to a later study.

To find the shadowed areas, we used the resampled multispectral Quickbird image obtained as described in the previous section. The four bands of this resampled multispectral image were then clustered into three clusters, using K-means clustering (MacQueen, 1967), where the cluster with the lowest mean value was selected as the shadow mask.

3.1.3. Segmentation of vehicles

For the segmentation of vehicles, we have, for reasons described above, used only the panchromatic image. The segmentation approach assumes that a definition of the road network to be analysed is available through a mask delineating this area. This can be derived from GIS data. The masks obtained through the analysis of the multispectral images are used to mask out areas within the road network corresponding to vegetation and dark shadows. For the area within the mask, the approach then assumes that the road surface is the dominating region covering the largest area within the mask. Based on this assumption, the histogram for the pixel values within the road network is analysed and the mean value of the road surface is determined at the peak of this histogram. Then Otsu's method for threshold selection is applied twice, once for the interval below this peak, and once for the interval above.

This approach results in a segmentation of the pixels into three categories: dark objects, road surface and bright objects. From this segmentation result, the connected components (connected pixels that have been given the same label) corresponding to the dark and the bright objects can then be identified. These components will then correspond to vehicles and to other similar objects such as smaller shadows, road markings etc.

3.2. Feature extraction

Segmentation will result in detection of different types of objects, where the objects corresponding to vehicles need to be identified. For this, we will use a classification approach, where specific characteristics (features) of the objects are first extracted from each object before the objects are classified based on these features. Two types of features have been used; *spatial* features describing the shape of the objects and *grey level* features describing properties related to the pixel values of the objects.

3.2.1. Spatial features

Spatial features were selected to be able to distinguish vehicles from other types of objects based on their shape. Vehicles are expected to be compact objects with a rectangular shape that have a width and a length within a certain size range and an orientation parallel with the road. Hence, different features that were expected to reflect these properties were tested. To select the initial set of these features we have built on previous experience from other projects on object recognition (see for instance Solberg and Solberg (1996)). The set of features that was selected is described below. Some of these features were only used in pre-classification, while others were only used for final statistical classification.

The following shape properties of a *region* were computed:

- *Area*: The number of pixels of the region.
- *Compactness*: $(\text{perimeter})^2/\text{area}$. Says how closely packed a shape is, where the most compact shape is a circle.
- *Angle deviation*: The difference between the direction of the road and the orientation of the object, where the orientation is computed from the central moments. This deviation is set to zero for objects that are close to quadratic, and do not have a clear direction.
- *Spatial spread*: The first invariant moment measures the spatial spread of the pixels relative to the size of the region.
- *Hu moments* (Hu, 1962): The Hu moments are a classical and commonly used set of features that can be derived from a shape. They are derived from the normalized central moments η_{pq} to provide a set of scale, position and rotation invariant features for pattern recognition. The first two Hu moments have been chosen as features. These are combinations of the normalized central moments η_{02}, η_{20} and η_{11}.
- *Height and width* of the bounding box of the rotated region. Here the bounding box is found by first computing the orientation of the object from the central moments, rotating the object to be aligned with the x-axis and finding the circumscribing rectangle of this rotated object. The height and width of this box will for vehicles correspond to the width and the length of the vehicle.
- *Elongation*: The ratio between the height and the width of the bounding box (computed as described above).
- *Rectangularity*: The ratio between the area of the region and the area of its bounding box (computed from the rotated region as described above). This says how rectangular a shape is (whether it fills out its bounding box).

3.2.2. Grey level features

In addition to the spatial features, features based on the grey level information (the pixel values) are also extracted. Vehicles are expected to appear as small objects contrasted by the background. Dark vehicles will appear as objects darker than the background and bright vehicles appear as objects brighter than the background. Bright vehicles will generally have a higher contrast to the background than dark vehicles.

Based on these considerations, features have been selected that say something about the outline and the contrast to the background in addition to the grey levels of the object itself:

- *Region mean*: The mean of the pixel values within the region.
- *Region standard deviation*: The standard deviation of the pixel values within the region.
- *Region gradient mean*: The mean of the gradients within the region.
- *Boundary gradient*: The mean gradient along the boundary of the region.
- *Local contrast*: The difference between the mean value of the region and the mean value of the local background.
- *Smoothness contrast ratio*: The ratio of the mean gradient magnitude (calculated within an extended bounding box of the region) to the mean gradient magnitude within the inner region area.

3.3. Classification

Classification is performed on the set of objects resulting from segmentation, and is performed in two steps. First a hierarchical, rule-based classifier is used to rule out as many non-vehicle objects as possible. The resulting potential vehicles are then further analysed through a final step of statistical classification. More details on this process are given in the next sections.

3.3.1. Rule-based classifier

The purpose of this initial, rule-based classification is to distinguish potential vehicle objects from obvious non-vehicles. The reason for doing this is that the class of objects that are non-vehicles will be very heterogeneous, containing objects with a large variation in shape and grey level signatures. The objects may represent road markings, shadows, reflections or parts of houses or other structures along the road. Hence, it is difficult to model this class of clutter objects in a meaningful way. On the other hand, some of these objects may be very different from vehicle objects, and may therefore easily be distinguished from the vehicles only by looking at one feature, or a small number of features.

A hierarchical, rule-based classifier following a coarse-to-fine strategy is defined. In work on coarse-to-fine classification, attributes are employed to recursively partition the set of hypotheses into ever finer and more homogeneous subsets (Gangaputra and Geman, 2006). In our study, the hierarchy was manually designed, selecting one feature at a time as the basis for splitting.

At the highest level, we separate between large and small objects based on the length of the object. We do this to separate large-sized vehicles from average-sized vehicles. Objects representing large-sized vehicles will, in general, have a more prominent appearance and well-defined shape than objects representing smaller vehicles. Hence, the large-sized vehicles are often more easily separated from clutter objects.

After the first partition, based on object length, we separate different types of clutter objects from potential vehicles, where features and the order of the tests are different for large and small objects. Features that are used in these tests include different geometric features like width, elongation, rectangularity, compactness and angle deviation. The test criteria at each level are set to avoid discarding potential vehicle objects. However, this cannot be completely ruled out for cases where segmentation has failed and resulted in fragmented or connected objects.

As a result of pre-classification, segmented objects will be partitioned into four main groups:

- Potential large-sized vehicles
- Large-sized clutter
- Small-sized clutter
- Potential normal-sized vehicles.

Partitioning into these specific groups is due to the hierarchical classification scheme that is used. The objects classified as potential vehicles in this process, are then further analysed through a last step of statistical classification while non-vehicle objects (large- and small-sized clutter) are discarded.

3.3.2. Statistical classifier

The purpose of statistical classification is to classify the remaining set of unclassified objects resulting from pre-classification. For this, we considered different approaches, based on linear and quadratic discriminant analysis.

Linear discriminant analysis (LDA) is a classical statistical approach for classifying samples of unknown classes, based on training samples with known classes which maximizes the ratio *between-class variance/within-class variance*. The decision boundaries of the LDA classification will be linear. For more details see Venables and Ripley (2002).

For QDA (Quadratic Discriminant Analysis) the decision boundaries will be quadratic. In general, QDA is based on class-dependent means and covariance matrices. Under the assumption that each class c is multivariate normal with mean μ_c and covariance matrix Σ_c, a sample x with unknown class is classified to the class c that minimizes

$$Q_c = (x - \mu_c)\Sigma_c^{-1}(x - \mu_c)^{\mathrm{T}} + \log|\Sigma_c| - 2\log\pi_c,$$

where the π_c's are the prior probabilities of the classes (Venables and Ripley, 2002). The Σ_c's and μ_c's are estimated from a training set consisting of samples with known classes.

QDA is more flexible than LDA, but more parameters need to be estimated. If n is the number of features, the number of parameters

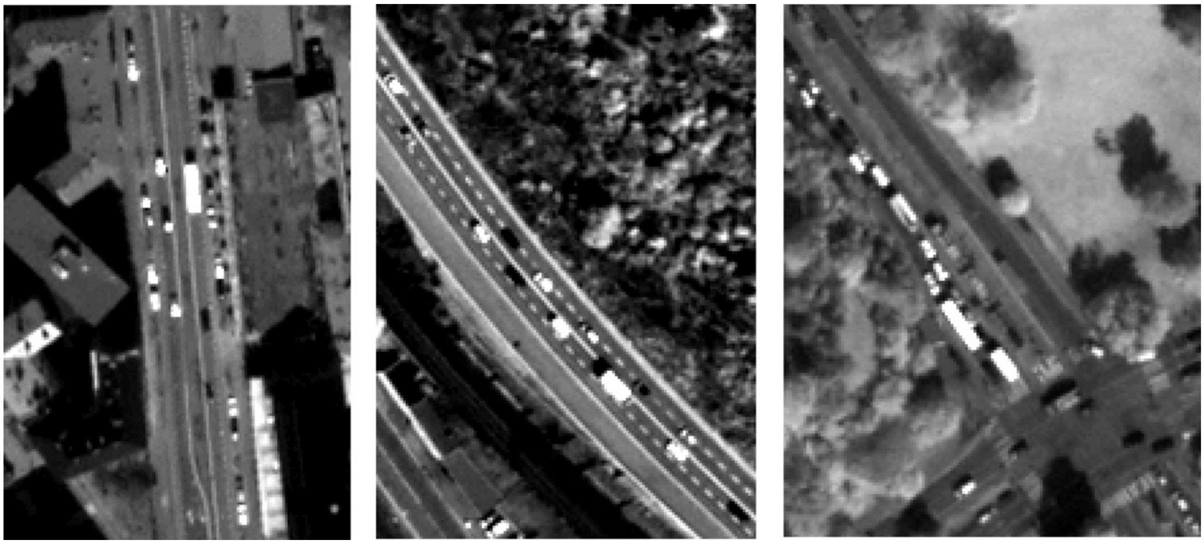

Fig. 1. Examples from smaller parts of each of the three areas. Left: GO, Middle: MV, Right: UV.

to be estimated is $n(n+3)/2$ for each class. For instance, with ten features, the number of parameters is 65. If the number of training samples is small, and typically less than the number of parameters involved in QDA, the estimates of the parameters based on the training set cannot be reliable. The high number of parameters to be estimated is mainly related to the covariance matrices. Common methods for coping with this problem are to constrain the covariance matrices, e.g. to use diagonal covariance matrices, or to use pooling, i.e. to estimate only one covariance matrix. If we for example assume that the features are uncorrelated, the covariance matrices are diagonal, and only $2n$ parameters need to be estimated for each class.

4. Experiments and results

4.1. Data set

Our dataset consisted of two Quickbird images over Oslo. One image was acquired early in May 2003 and the other late in May 2006. Both images are acquired around 11 in the morning, with a sun elevation angle of 46° in early May and 51° in late May. Three areas were selected for analysis, representing different types of roads and traffic situations:

- GO (Gamle Oslo): Inner-city roads, speed limit 50 km/h.
- MV (Mosseveien): Main road to/from Oslo, speed limit 80 km/h.
- UV (Ullevål): Inner-city roads, speed limit 50 km/h.

Examples from smaller parts of each of these three areas are shown in Fig. 1.

Subimages covering these three areas were extracted from each of the two images. This resulted in a set of 6 images, where 3 of these sub-images, one from each area, were used for training while the other 3 were used for test. For each image, a road mask was constructed manually.

To train the classifier and to verify the results, 'ground truth' was established through manual analysis and classification of the objects. However, even manual classification is difficult at this resolution. Training was therefore only based on objects where the interpretation was unambiguous.

For *verification of the classification results*, it is necessary to determine the class of all the objects. To be able to reflect at least some of the ambiguity in the manual interpretation of the satellite images, we have produced two independent manual classifications performed by two different persons. The results of the automatic analysis can then be compared with both the manual classifications.

Two different types of manual counts have been produced to evaluate the detection and classification results:

- A *manual count* containing the positions of all vehicles within the road network that each person has observed. This has been obtained through manual inspection of the panchromatic image. The purpose of these counts is to use them for comparison with the *segmentation* results.
- A *manual classification,* where the set of objects sent to the classification has been manually labeled with the four classes: dark clutter, bright clutter, dark vehicle, bright vehicle. The purpose of these manual classifications is to use them for comparison with the results from the *automatic classification*.

4.2. Experiments

4.2.1. Segmentation

Segmentation was performed on the sub-images from both the training set and the test set. First, vegetation and dark shadows were identified based on the multispectral images. A road mask was then created from a vector representation of the road network to indicate the area to be further analysed. Then, the panchromatic images were analysed within the road mask for regions not covered by vegetation or dark shadows.

4.2.2. Preclassification

The preclassification was performed on all the connected components identified as dark or bright objects in the segmentation phase.

4.2.3. Training

For training, the objects resulting from the segmentation and pre-classification of the three training images were manually analysed. These objects were manually classified into one of four classes:

- Dark clutter
- Bright clutter
- Dark vehicle
- Bright vehicle.

Objects for which the class was difficult to determine were not included in the training set. For all but the class 'Bright clutter', between 140 and 190 training samples were obtained; (Dark clutter: 144, Bright clutter: 38, Dark vehicle: 158, Bright vehicle: 186). A feature vector of dimension 14, consisting of a selection of spatial and grey level features, as described in Section 3.2, was used

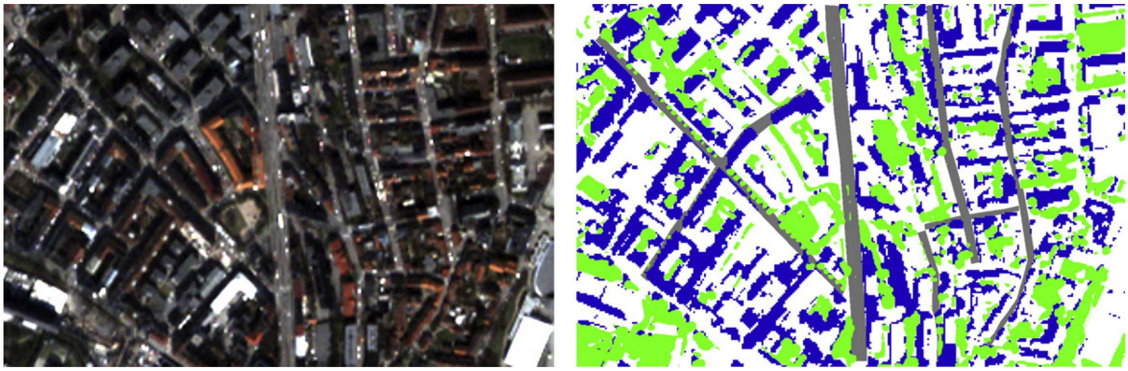

Fig. 2. RGB image (left) and the combined mask (right) with a subset of roads (grey), vegetation (green) and shadows (blue).

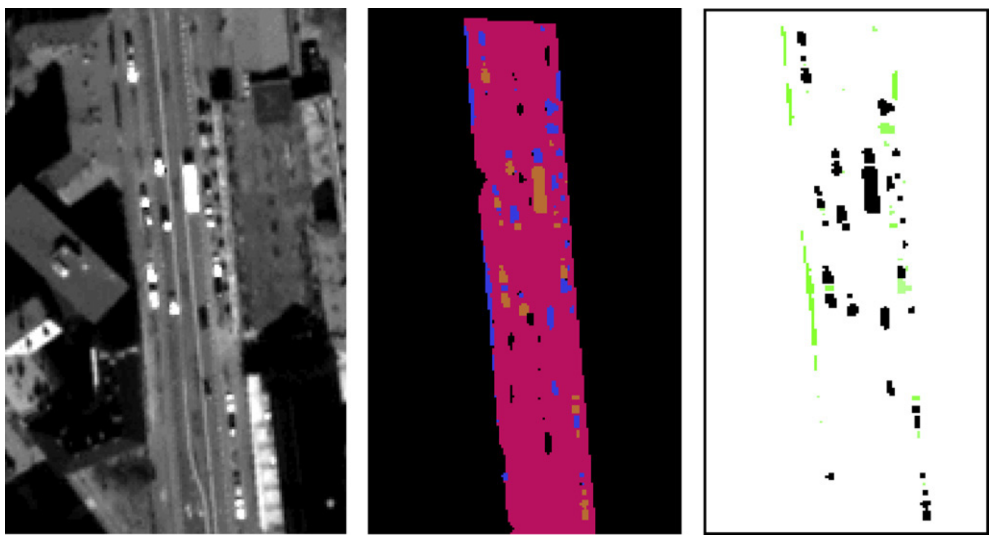

Fig. 3. Result of the segmentation on a part of the road from the GO subimage. Left: Grey level image, Middle: Segmented road surface (pink), dark objects (blue), bright objects (orange) and not within mask (black). Right: Result of pre-classification, where the regions that are black correspond to objects identified as potential vehicles.

to represent each object. The selected set consisted of the following features: *length, width, compactness, elongation, rectangularity, boundary gradient, spatial spread, contrast, smoothness, region mean, gradient mean, variance* and *Hu moments*.

The feature selection was made based on data analysis and tests on the training set. This was performed by analysing the distribution of features values for the different classes of objects, inspecting scatter plots for these classes for different combinations of features and performing preliminary classifications on parts of the training set.

4.2.4. Classification

LDA, general QDA and QDA with diagonal covariance matrices were tested for classification. The performance of the general QDA and the QDA with diagonal covariance matrices was quite similar, while the LDA did not demonstrate a comparable performance. The three most common classes had enough samples to accommodate the general QDA, and the results reported later are from the use of this classifier. All the three test images were classified using this approach.

4.3. Results and discussion

4.3.1. Segmentation and preclassification

In the segmentation process, vegetation and dark shadows are first identified. A result of this process for the sub-image GO, can be seen in Fig. 2.

The further segmentation is then performed on the areas within the road network that are not covered by vegetation or dark shadows. This part of the road network is segmented into road surface, dark objects and bright objects. A following pre-classification is then used to filter out segmented objects that are not potential vehicles. A result of this process for a part of the sub-image GO can be seen in Fig. 3.

In Table 1 the results from segmentation and pre-classification are summarized. The results are compared with counts obtained by two independent manual counts based on the same image. The total manual vehicle count given in the first three columns of Table 1, gives the results obtained by the two independent manual counts performed by two different persons (P1: person 1, P2: person 2). The column marked consensus gives the number of objects (positions) where both persons agree that there is a vehicle. The next columns summarize the number of vehicles that were missed by the segmentation and pre-classification process compared with the two manual counts.

As can be seen from the counts, there is a number of vehicles that are missed in this process. According to the two manual counts, 69–73 vehicles are lost in this process. Inspection of the images for these cases, revealed that this is generally due to bad contrast, where the vehicle object is hardly visible at all or severely fragmented. In the segmentation process, this means that these vehicles cannot be separated from the background, or that only very small parts of the objects that do not have the shape of a vehicle are found. Examples illustrating this are given in Fig. 4. As can be seen here, the vehicles often consist of several parts; a

Table 1
Result of segmentation and preclassification

Image ID and total number of segmented objects per image	Total **manual** vehicle count			Vehicles missed before classification according to manual counts	
	P1	P2	Consensus	P1	P2
Image GO2 (528)	154	160	144	26	20
Image MV1 (255)	96	93	91	26	21
Image UV1 (836)	146	134	127	31	28
Total	396	387	362	73	69

Table 2
Result of manual classification

	Total number of objects classified	Number of objects manually classified as vehicles			Percentage overlap between P1 and P2 (%)
		P1	P2	Consensus	
Image GO2	205	126	126	106	80.5
Image MV1	83	59	68	57	84.9
Image UV1	211	111	111	94	84.3
Total	499	296	305	257	83

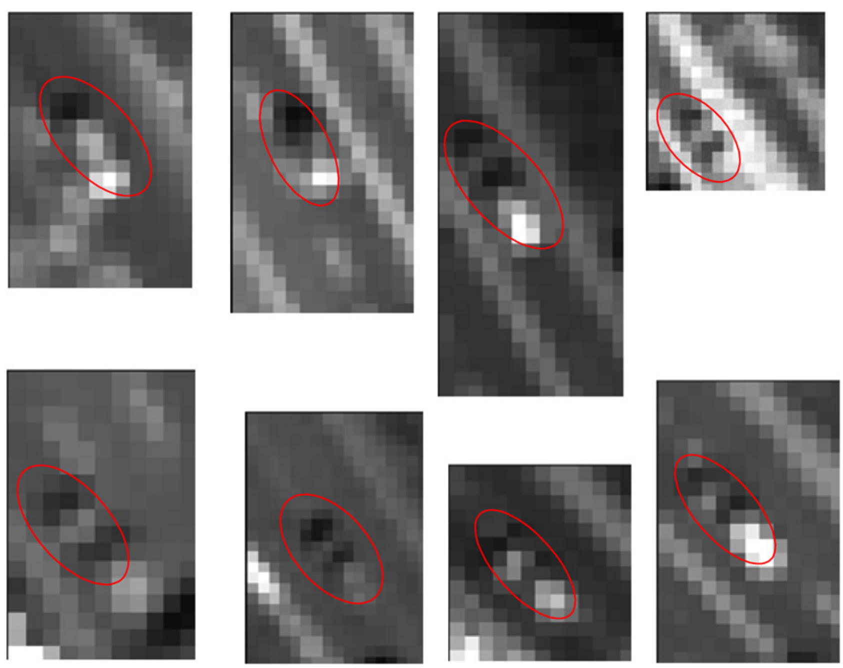

Fig. 4. Examples of manually identified vehicle objects that are not properly segmented. These objects have low contrast with the background and are therefore fragmented into several very small parts that are not sufficiently large to be considered as a potential vehicle.

shadow, the vehicle and a reflection from the rear window. This may indicate that the sun angle has an effect on how the vehicles are imaged.

Looking at the results for the three areas separately, the highest percentage of missed objects appears for the MV sub-image. This area also contains the roads with the highest speed limit of the three test areas. Hence, this may indicate that the high speed is a factor that affects the contrast conditions in this case.

4.3.2. Classification

Table 2 summarizes the results of the two independent manual classifications that were performed. This classification was performed on the objects remaining after the pre-classification. The column to the left gives the number of objects that were classified. The next three columns give the number of these objects that were classified as vehicles by person 1 (P1) and person 2 (P2), while the column marked consensus gives the number of objects where both manual classifications agreed that the object in question was a vehicle. The rightmost column gives the percentage of objects where both manual classifications agreed on the class (vehicle/no-vehicle). In total the two independent manual classifications agree on the classification for 83% of the objects. This indicates that the interpretation of these images is not straightforward.

Table 3 summarizes the results of automatic classification. The column to the left gives the total number of objects that were classified as vehicles. The total here is 303 vehicles, which is very close to that of the two manual classifications that gave 296 and 305 vehicles respectively (as reported in Table 2). The next two columns give the number of objects that were correctly classified as vehicles according to the two manual classifications. Again the numbers obtained here (256 and 252) are comparable with the number of vehicle-classifications where the two manual classifications agree (257) from Table 2.

The two rightmost columns of Table 3 give the correct classification rates obtained as compared with the two manual classifications, P1 and P2. For P1 the total rate is 82.8%, while for P2 the total rate is 79.2%. These rates are close to the percentage of

Table 3
Result of automatic classification

QDA	Objects classified as vehicles	Correctly classified as vehicles		Correct classification rates	
		P1	P2	P1 (%)	P2 (%)
Image GO2	135	112	112	82.0	82.0
Image MV1	65	56	59	86.7	81.9
Image UV1	103	88	81	82.0	75.4
Total	303	256	252	82.8	79.2

objects for which the two manual classifications are in agreement (83%). Hence, this indicates that the performance of the automatic classification is comparable with that of a manual classification.

As can be seen from the manual counts, obtaining ground truth for these images is difficult. It could therefore be useful to investigate additional types of ground truth. One possible solution could be to acquire counts from inductive loop counters, where such equipment is installed. From these, it is possible to obtain the number of vehicles around the time of image acquisition. By using additional information on vehicle speed, it should then be possible to estimate the expected number of vehicles along a specific stretch of road at the time the satellite image was acquired. This can then be used as an additional and independent source for evaluation of the number of vehicles detected along the same stretch of road. This approach was not possible in this study, as no such counters were installed along the roads in question. However, this will be considered in future studies.

4.3.3. Overall performance

The results have shown that even manual classification is difficult at this resolution, as the appearance of the observed objects can be quite ambiguous. This indicates that the resolution is close to the limit of what makes vehicle detection possible. Still, automatic classification displays the same performance as a human performing a classification of the objects, which is a very good result.

However, in the initial segmentation process, vehicle objects are lost, due to low contrast and fragmentation. It seems that the lower sun angle at these latitudes (60° north) complicates this matter, giving rise to more and longer shadows and more reflections. This can make vehicles appear as several blobs of shadow, vehicle and reflections. At this resolution, each of these blobs can be very small, only 1–3 pixels, and therefore difficult to identify as meaningful objects. In future work, approaches that are able to recognize and combine even small blobs resulting from shadows and reflections with those resulting from vehicles may be considered to help the vehicle detection. This may, however, require quite complex models to be able to cover for different light angles and vehicle orientations. In cases where shadow objects are larger (typically alongside a vehicle, and not behind or in front of) the shadows may be used to signal the existence of a vehicle or verify a detection. This will, however only be useful for some light angles relative to the vehicle position.

In the experiments reported here images from a region surrounding Oslo, acquired during spring time, have been used. The general approach should however be able to also work for other regions and for other seasons, provided that weather and light conditions allow for imagery to be acquired. Preliminary experiments (not included in this study) performed on images covering regions further north in Norway (latitudes varying from ca 63° to 70° north) also indicate this. However, future studies should investigate this in more detail.

In total, the number of detected vehicles (303) is approximately 20% less than that obtained through the manual counts (consensus: 362), and the underestimation is due to the objects lost during segmentation. It is difficult to compare our results with those obtained in other studies of vehicle detection from satellite images. The few studies that exist have concentrated on large highways, mainly in the US, where conditions are quite different.

For highway areas, Sharma et al. (2006) report quite good results for their method based on Bayesian background transformation. Here, the vehicle count is close to the ground truth (for a total of 160 vehicles) and the number of false positives and false negatives is small. This method does, however, require a background estimate, and such an estimate will seldom be available and is difficult to obtain automatically. Their next best method also obtains a good estimate of the number of the total vehicles. However, the false negative rate and the false positive rate are both in the proximity of 25%. These results are obtained for wide and open highways with very little clutter. They are obtained by comparison with manual counts, and no problems or ambiguities in the manual counts have been reported. This indicates that the vehicles in this study are quite easily distinguished from the road surface.

5. Summary and conclusion

In this paper, we have demonstrated a classification-based approach for vehicle detection in satellite images. The approach utilizes panchromatic information for vehicle classification and multispectral information for masking of vegetation and shadows. Classification of segmented objects is then performed in two steps, using a rule-based pre-classification and a final statistical classification. Tests have been performed on Quickbird images from two different years, taken over Oslo, Norway. The roads for which the vehicle detection has been tested consist of a mixture of inner-city roads and main roads.

The results of the segmentation and classification have been compared with manual counts and classifications. These manual classifications have been performed by two persons independently, and have shown that even manual classification is difficult at this resolution, as the appearance of the observed objects can be quite ambiguous. This indicates that for these roads with this resolution and under these conditions it is difficult to distinguish vehicles from other objects. However, comparisons with manual classifications have shown that the automatic classification is good, and that the results are very similar to those obtained through manual classification of the objects. The segmentation process is however more of a problem. Approximately 20% of the vehicles are lost in this process due to low contrast and fragmented objects.

In conclusion, the classification results that have been obtained are very promising. Hence, this seems to be a good approach for inner-city roads with different types of clutter. However, in future studies we want to look more into the problems of the segmentation process and investigate the possibility of improving the performance of this process. At the same time, we think it is necessary to compare the vehicle detection results with other types of traffic information, resulting for instance from inductive loops.

There are also some additional challenges for a satellite-based system. For instance, high buildings and their shadows as well as vegetation can occlude parts of the road network. Satellite-based traffic data can then not be derived for these parts. Since traffic load often varies relatively smoothly, statistics from nearby areas that are observable may still provide useful insight into the situation for these occluded areas.

Another challenge is the availability of satellite images. For the geographical regions used in this study we would, due to weather conditions, expect to acquire 1–2 images of sufficient quality per month, and further north, no images would be available during winter time due to light conditions. However, an approach like this is not intended to replace the current systems for continuous traffic surveillance, but as an additional and complementing source of information. Useful links may also be established between counts from traditional equipment, existing traffic estimates (annual average daily traffic — AADT) and the satellite-based information.

In future studies, we want to investigate more closely the consistency of the results and the level of performance that is needed for this type of satellite-based traffic counts to be useful, and how this information can be linked to other types of information. For areas where no information is available today, satellite-based counts may contribute with useful information, even with underestimation, and it can also give valuable insight into vehicle distribution patterns over the road network.

Acknowledgements

The work presented in this paper was carried out in projects financed by the Research Council of Norway and the European Space Agency (ESA/ESRIN).

References

Alba-Flores, R., 2005. Evaluation of the Use of High-Resolution Satellite Imagery in Transportation Applications. Technical Note, Department of Electrical and Computer Engineering, University of Minnesota Duluth, August.

Gangaputra, S., Geman, D., 2006. A design principle for coarse-to-fine classification. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, New York, June 17–22. pp. 1877–1884.

Gerhardinger, A., Ehrlich, D., Pesaresi, M., 2005. Vehicle detection from very high resolution satellite imagery for the development of a societal activity index. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3/W24), 83–88.

Hinz, S., 2004. Detection of vehicles and vehicle queues for road monitoring using high resolution aerial images. Photogrammetrie Fernerkundung Geoinformation 2004 (3), 201–213.

Hu, M-K., 1962. Visual pattern recognition by moment invariants. IRE Transactions on Information Theory IT-8, 179–187.

Leitloff, J., Hinz, S., Stilla, U., 2005. Vehicle queue detection in satellite images of urban areas. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 8/W27), on CD-ROM.

MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley. vol. 1. pp. 281–297.

McCord, M.R., Merry, C.J., Goel, P., 1998. Incorporating satellite imagery in traffic monitoring programs. In: North American Travel Monitoring Exhibition and Conference, North Carolina USA, May 11–15.

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

Schlosser, C., Reitberger, J., Hinz, S., 2003. Automatic car detection in high resolution urban scenes based on an adaptive 3D-model. In: Proceedings of 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, Berlin. pp. 167–171.

Sharma, G., 2002. Vehicle detection and classification in 1-m resolution imagery. Master of Science Thesis, Ohio State University.

Sharma, G., Merry, C.J., Goel, P., McCord, M.R., 2006. Vehicle detection in 1-m resolution satellite and airborne imagery. International Journal of Remote Sensing 27 (3–4), 779–797.

Solberg, A.H.S., Solberg, R., 1996. A large-scale evaluation of features for automatic detection of oil spills in ERS SAR images. In: Geoscience and Remote Sensing Symposium, IGARSS '96. Remote Sensing for a Sustainable Future 3, 1484–1486.

Stilla, U., Michaelsen, E., Soergel, U., Hinz, S., Ender, J., 2004. Airborne monitoring of vehicle activity in urban areas. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences 35 (Part B3), 973–979.

Toth, C.K., Grejner-Brzezinska, D., 2005. Traffic flow estimation from airborne imaging sensors: A performance analysis. In: Proceedings ISPRS Hannover Workshop on High-Resolution Earth Imaging for Geospatial Information, Hannover, Germany, May 17–20. http://www.isprs.org/publications/related/hannover05/paper/papers.htm (accessed 09.09.2008).

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, fourth edition. Springer.

Zhao, T., Nevatia, R., 2001. Car detection in low resolution aerial images. In: Proceedings of International Conference on Computer Vision, Vancouver, Canada July 9–12, 2001. pp. 710–717.