Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

# SatTrafikk

## Vehicle Detection in Satellite Images for Development of Traffic Statistics

## Project results 2007

Note no SAMBA/18/08

Authors Siri Øyen Larsen, Hans Koren, and Rune Solberg

Date May 2008

## Norsk Regnesentral
Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modeling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

## Norsk Romsenter
Norsk Romsenter (Norwegian Space Centre, NSC) is a government agency under the Ministry of Trade and Industry. NSC promotes the development, co-ordination and evaluation of national space activities as well as supports Norwegian interests in the European Space Agency (ESA). Earth observation involves all activities related to collection of information on the Earth's surface or atmosphere from instruments on board satellites. The Norwegian Space Centre's application programme supports users, research communities and businesses in testing the potential of Earth observation from satellites. Priority is given to the development of applications having public benefit.

## Statens Vegvesen
Statens Vegvesen (The Norwegian Public Roads Administration (NPRA)) is responsible for the planning, construction and operation of the national and county road networks, vehicle inspection and requirements, driver training and licensing. It is also authorized to grant subsidies for ferry operations. The objective of the NPRA is to develop and maintain a safe, ecofriendly and efficient transport system. This is being done on a sound, professional basis by interacting with politicians, users and other interested parties.

**Norsk Regnesentral**
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

**Besøksadresse**
*Office address*
Gaustadalléen 23
NO-0373 Oslo, Norway

**Telefon** · telephone
(+47) 22 85 25 00
**Telefaks** · telefax
(+47) 22 69 76 60

**Internett** · internet
www.nr.no
**E-post** · e-mail
nr@nr.no

| | |
|---|---|
| **Title** | **SatTrafikk - Vehicle Detection in Satellite Images for Development of Traffic Statistics – Project results 2007** |


| | |
|---|---|
| **Authors** | **Siri Øyen Larsen, NR** |
| | **Hans Koren, NR** |
| | **Rune Solberg, NR** |



| | |
|---|---|
| Date | May |
| Year | 2008 |
| Publication number | SAMBA/18/08 |



## Abstract

The road network is a key resource of enormous value for the nation. A satellite image is covering a large area instantaneously and can thereby be a source of road traffic information in a snapshot of time. Manual vehicle counting is not realistic, but automatic image analysis methodology based on pattern recognition is a promising alternative. The "Road Traffic Snapshot" project (2006-2007) showed that such a future system for automated road traffic counts is feasible.  The follow-on project SatTrafikk has improved the methodology and verified it on a far larger set of satellite images.

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 was implemented and tested against manual vehicle counts. We also compared the results to traffic statistics estimated from single-point measurements. The majority (90%) of vehicles that are found in the segmentation step are correctly classified as vehicles. Comparing the number of cars found in the image to the number of cars that according to single-point measurements are expected to be found in the image, the manual counts had accuracy between 79 % and 120%, while the automatic counts had accuracy between 68% and 157%. Some shadows and road marks are hard to distinguish from vehicles. Some vehicles have low contrast and are not captured in the segmentation step.



| | |
|---|---|
| Keywords | Remote sensing, pattern recognition, vehicle detection, road traffic statistics, very-high-resolution satellite images, QuickBird |
| Target group | Road traffic authorities |
| Availability | Open |
| Project number | 220 339 |
| Research field | Earth Observation |
| Number of pages | 45 |

# Contents

# Executive summary

The road network is a key resource of enormous value for the nation. The current primary source of traffic statistics is measurement stations based on induction loops counting vehicles that pass a given point in the road system over time. In cases were the traffic distribution in an entire road network is of interest, this methodology has evident shortcomings due to the very limited geographical coverage of such a system. A satellite image is covering a large area instantaneously and can thereby be a source of road traffic information in a snapshot of time. Manual vehicle counting is not realistic, but automatic image analysis methodology based on pattern recognition is a promising alternative. These ideas were tested out in the ESA project “Road Traffic Snapshot” (RTS) in 2006-2007. The project showed that a future system for automatic road traffic counts is feasible. Based on this positive outcome an initiative was made to continue the work in the project described in this report. The outcome of the first year of the SatTrafikk project is described in this report.

The approach that was outlined in the RTS project has been used as a basis for the work that is presented in this report. We suggest an approach with the following steps. First, we make a road mask by manual delineation of the roads in the image. We also make a vegetation mask, which is necessary in cases where the crowns of trees by the road cover parts of it, or vegetation has been planted between the two lanes of the road. The vegetation mask is computed using the multispectral image. After masking, the rest of the algorithm focuses on the masked part of the panchromatic image. Segmentation locates objects of dark or bright intensity. The segmentation routine exploits the information that is given in the image histogram. The method that was used in RTS has been significantly modified. In stead of using one global threshold for each stage of the segmentation (segmentation of dark and bright segments each consitute a stage), we use a strict and a loose threshold, and combine the results. The dark segments are then checked against a road edge mask; dark segments that are located close to the edge of the road are assumed to be shadows, hence, they are rejected. The latter routine replaces the shadow mask that was used in RTS. The next step is feature extraction. The segments are preclassified according to the standard features area, elongation, length of bounding box, region intensity mean, and region gradient mean. Segments that obviously do not fit with the expected attributes of a vehicle are rejected. The remaining segments are sent to the main classification routine, together with the extracted feature values; region intensity mean, region gradient mean, region intensity standard deviation, length, 1st Hu moment, spatial spread, and “distance to nearest vehicle shadow”. The first six features are used for statistical classification, using the maximum likelihood method. The last feature is helpful as contextual information and is used for a postclassification step. In this work we have developed an approach for estimation of a vehicle shadow mask, which is used to compute the distance to nearest shadow for each segment. The idea of using a rule-based preclassification followed by a maximum likelihood classification was presented in the RTS project. In the SatTrafikk project we have made a thorough study of features, re-evaluated the preclassification parameters, introduced the use of class prior probabilities, and altered the class definition.

All the routines described above have been implemented as updates to the “Road Traffic” software tool that was developed in the RTS project. Methods for training and editing segment labels have also been implemented. The training step is used to teach the classifier how to differentiate between different classes of segments. The user is guided through a set of extracted

segments and asked to label them. Based on a set of labeled segments, the class statistics are calculated. The edit step is performed after classification, and the user is led through the segments in the image and asked to either change or accept the given labels.

The editing of segments after classification has no chance of finding vehicles that may have been missed by the segmentation step. Thus, other types of validation activities have also been performed. We have compared the results of the automatic vehicle detection to the results of manually counting vehicles in the images. Furthermore, results have been assessed based on traffic statistics from in-road counting stations. The majority of vehicles that are found in the segmentation step (90%) are correctly classified as vehicles. According to manual vehicle counts around 70% of the segments that were classified as vehicles were correctly classified. Compared with the number of vehicles that are expected to be present in the image when using estimates from in-road equipment, the automatic count performs overestimation in two out of six images, by 110% and 157% respectively. This is mainly due to tree shadows that enter far into the road, as well as high contrast road marks, which tend to be confused as vehicles. In the remaining images there is an underesimation of vehicles, by 68% in three images and 81% in one image. The segmentation routine should be improved so that it finds even vehicles with low contrast.

# 1 Introduction

## 1.1 Background

The road network is a key resource of enormous value for the nation. Maintenance and development of the road network is an important issue of several public entities, including the Norwegian Public Roads Administration (Statens Vegvesen Vegdirektoratet, SVV). Traffic statistics is necessary information for these purposes. The current primary source of traffic statistics is measurement stations based on induction loops counting vehicles that pass a given point in the road system over time. A number of statistical measures might be derived from these data. The most important information is the Annual Day Traffic (ADT), a measure of the traffic distribution during a day at a specific point and averaged over a year. In Norway, ADT is currently derived from point measurements and statistical tools developed by NR.

In cases were the traffic distribution in an entire road network is of interest, 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 counters is far from realistic and alternatives are necessary. A potential alternative is traffic counting in very-high-resolution satellite images, like QuickBird (0.6 m pixel resolution). A satellite image is covering a large area instantaneously and can thereby be a source of road traffic information in a snapshot of time. 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.

Motivated by the abovementioned challenges and ideas, NR carried out in 2006-2007 the project "Road Traffic Snapshot" in collaboration with the SVV and the Institute of Transport Economics (Transportøkonomisk instititt, TØI). The project was funded by the European Space Agency (ESA), and its main objective was to develop and test the necessary methodology for vehicle detection in very-high-resolution satellite imagery, like QuickBird images. The content of the project was to: (i) Define user requirements and technical specifications for the service; (ii) Make a service case implementation and; (iii) Validate and evaluate the implemented methodology, and discuss possibilities for the evolution of the project.

The Road Traffic Snapshot project showed that a future system for automatic road traffic counts is feasible. Even if such a system observes only a snapshot of the road traffic situation, the results indicated that this information is relevant not only for that particular instance in time but also relevant for estimating the traffic load in a longer time period. For a full description of the project, see [1].

Based on the positive outcome of the Road Traffic Snapshot project, SVV and NR made an initiative to continue the work. The Norwegian Space Centre (NSC) supported the idea and contributed with funding together with SVV. The outcome of the first year of this project, SatTrafikk, is described in this report.

## 1.2 Project objectives

The objectives of the first year of the SatTrafikk project have been to:

1. Validate the methodology developed in the Road Traffic Snapshot project on a far larger satellite image dataset

2. Improve and optimise the methodology based on the experience of the enlarged dataset

If the project is successful with respect to these two objectives, the intention is to develop the necessary methodology to estimate ADT from the vehicle detection results by:

3. Developing a statistical model for estimation of the speed of the vehicle based on local speed regulations and the distance between the vehicles as observed in the satellite image

4. Developing a statistical model for estimation of the number of vehicles crossing a specific point in the road based on 2 and 3 above.

The work towards objectives 1-2 was carried out in 2007-2008, and the work towards objectives 3-4 is planned carried out in 2008-2009.

## 1.3 About this report

This report has the following structure. In chapter 2 we explain the methods that concern vehicle detection. These include segmentation, feature extraction, classification, and detection of shadows. In chapter 3 we describe the image data and the set of roads that were selected for inspection. The production of road masks is also treated in this chapter. The validation activities and results are presented in chapter 4. Finally, in chapter 5, we reach some conclusions and make recommendations for future work.

# 2 Methods

In the ESA project "Road Traffic Snapshot" [1] different methods for vehicle detection were tested and suggestions were made regarding which approach to use. In our recent study, the methods have been improved, and some new algorithmic ingredients have been added. In the current chapter, we will walk through each step of the vehicle detection, describe the current status, and discuss the problems that were discovered, as well as the sougth improvements.

The detection methods have been developed using the image data from Kristiansund, Bodø, and Sollihøgda, where manual road masks have been made (section 3.2.1).

## 2.1 Masking

The search for vehicles was restricted to the parts of the image that contain roads. Hence, a road mask and a vegetation mask were applied to the panchromatic image before further processing took place. The shadow mask suggested in [1] has been discarded in this stage of the algorithm. The making of road masks will be discussed in section 3.2.

### 2.1.1 Vegetation mask

Trees shadowing the road can be a problem for vehicle detection. There is little to do about the problem of vehicles hidden underneath tree crowns etc. However, it is wishful to remove trees (and other vegetation) that partly covers the road before searching for vehicles on the road. We have used a vegetation mask, derived from multispectral information from the same scene. The NDVI (normalized difference vegetation index) is computed from the multi spectral image, after resampling to the resolution of the panchromatic image, using cubic interpolation. We then find the appropriate threshold from application of Otsu's algorithm to the resulting NDVI image, and use this to produce a vegetation mask. A detailed description of this method can be found in [1].

### 2.1.2 Shadow mask

In [1] we suggested to apply a shadow mask in order to remove shadows from tall buildings, trees near the road, etc., as a part of the preprocessing step. An algorithm using (K-means) clustering of the multispectral image was developed. This method did not appear satisfactory when applied to the image data of the current study. An explanation may be that the data in the first study was mainly inner city roads, while the latter project involves roads from less populated urban areas, as well as some countryside highways. The K-means clustering divides the image into three (or four) clusters, based on the spectral values in the four bands of the multispectral image. Areas that posess similar spectral values are assigned the same cluster label. The darkest cluster, i.e., the cluster with the lowest mean value, (or the two darkest areas, in the case that four clusters were used), is assumed to represent shadows. The main problem about this approach is that it depends strongly on the global properties of the image, and what type of areas that are contained in the image. Due to the large amounts of data in one satellite image, and for practical reasons, the original image is divided into subimages before processing. This means that a specific geographic location may be assigned different cluster labels depending on what subimage it happens to be a part of.

In the current study we have taken a different approach in order to remove shadows caused by vegetation close to the road. The shadow mask is completely removed from the preprocessing routine. Instead, a road edge mask is used to remove dark segments *after* the main segmentation routine (section 2.2.2.1).

## 2.2 Segmentation

The main segmentation routine is based on finding segments that are darker or brighter than their surroundings. The routine is applied to the masked panchromatic image, i.e., we only look for segments (potential vehicles) on the road.

### 2.2.1 The main segmentation method

We have developed an improved version of the previously used algorithm. The approach is based on separating the road pixels into three intensity intervals; low intensity - representing dark objects on the road, medium intenisty – representing the road cover (asphalt), and high intensity – representing bright objects on the road. Plotting image histograms of the test data we observe that there are typically two modes in the histogram; a small peak in the lower part of the intensity scale, followed by a much larger peak, see Figure 2.1. The histograms all have a long tail to the right of the main mode. The dominating class constitutes the road (asphalt) pixels. The high intensity tail corresponds to bright objects, agreeing with the fact that bright objects in the images appear in a wide range of intensities. (Some bright objects are only slightly brighter than the local background, while others produce very high intensity values). The lower (small) mode is representative for dark objects.

In [1] we suggested to use Otsu's method in two stages. Otsu's method decides upon a global threshold that best separates the two classes of gray tone levels in an input histogram. The suggested method first finds the mean intensity value. This value will be approximately at the

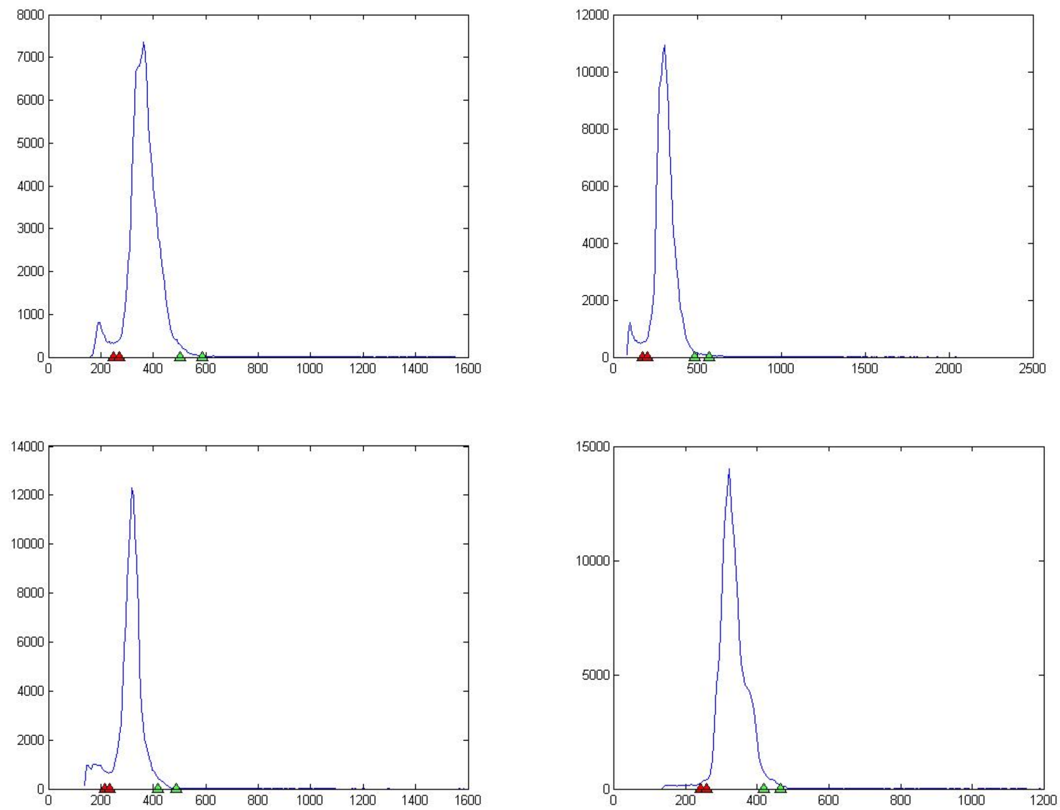



Figure 2.1 Image histograms for four test images. The histograms look relatively similar; there is a dominating class in the middle of the gray tone scale, a small dark class to the left, and a bright class that spans a number of high gray tone values. The red and green arrows indicate segmentation thresholds for the dark and bright class, respectively. The threshold closest to the peak is the weakest threshold for each class.

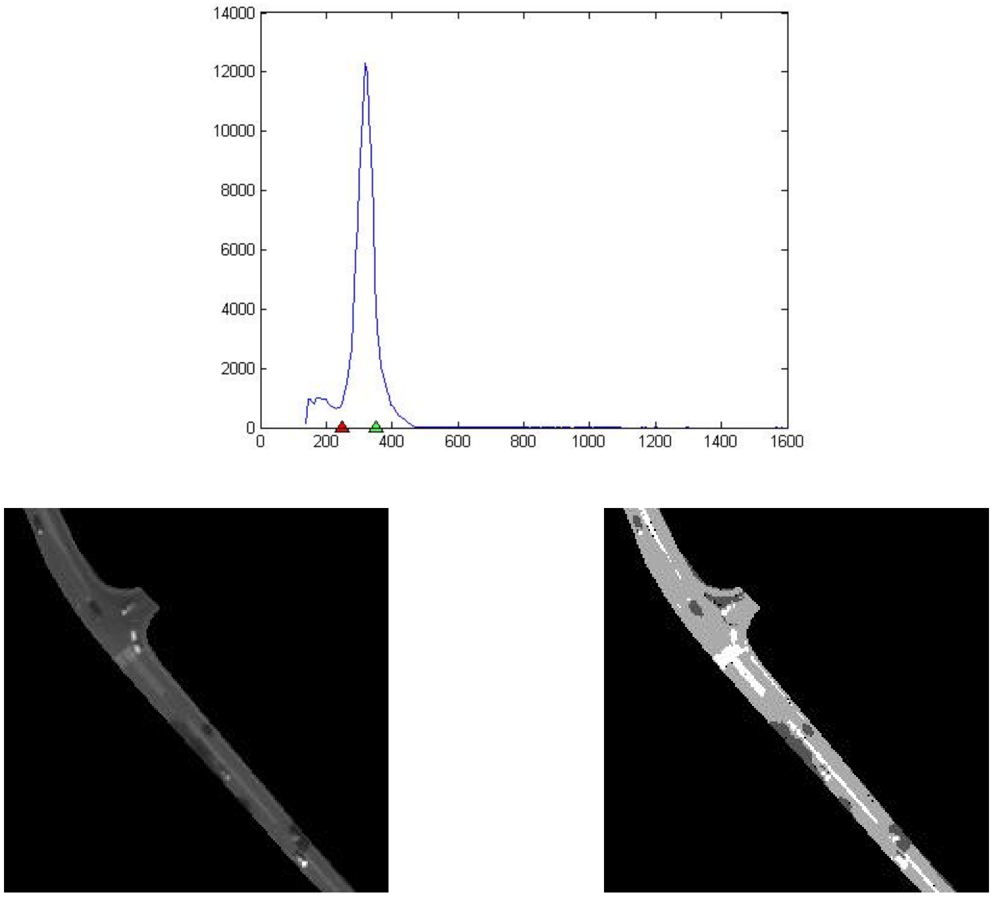



Figure 2.2 Estimating the correct intensity thresholds for segmentation is not easy. The left image is a subimage of the masked panchromatic image represented in the histogram. The right image shows the corresponding segmentation result, using the old segmentation approach, with the thresholds as marked in the histogram. As shown in the images, the thresholds do not provide the desired results. From the histogram we see that both thresholds (especially the upper) should be adjusted (moved away from the main peak).

main peak of the histogram. The threshold for the dark class is found by applying Otsu to the part of the histogram that resides to the left of the mean. All pixels below this threshold are assumed to belong to a dark object. Next, Otsu's method is applied to the right part of the histogram, and the resulting threshold is used as a lower limit for pixels in the bright class. New tests of this method have been performed, and indicate that the method lacks robustness. Figure 2.2 shows an example of where it fails. The lower threshold is too high, and the upper threshold is too low. Experiments showed that instead of splitting the histogram at the mean, before applying Otsu, it should be split a bit to either side of the mean (depending on which side we are seeking the threshold). It is not trivial to find the exact right threshold. A reason that the old method failed may be that it relies too heavily on correct estimation of the two global thresholds. In the current study we have developed a more flexible method; for each segmentation step (i.e., segmentation of dark and bright objects), two thresholds are found. (We may loosely speak of this as a strict and a weak threshold that constitue the limits of some sort of confidence interval for the correct threshold.) The thresholds are used for hysteresis thresholding, i.e., finding the segmentation results for the weak and the strict thresholds

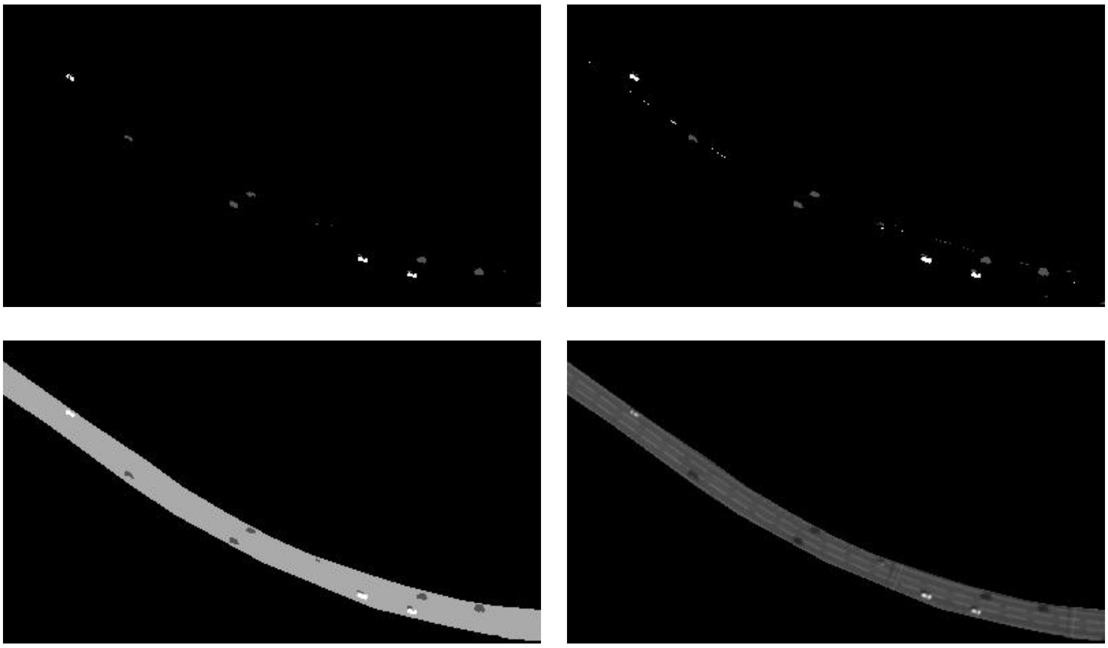


Figure 2.3 Upper left image shows the result of segmenting the image with the strict thresholds. White regions are above the strict threshold for bright regions, while dark gray regions are below the strict threshold for dark regions. In the upper right image we see the corresponding result using the weak thresholds. The segmented image below on the left is the final result after hysteresis thresholding. Light gray color indicate the road. On the right, the masked panchromatic image is shown.

separately, we combine them so that we keep only segments resulting from the weak threshold that contain segment(s) from the strict threshold. Using the strict threshold alone, there is a tendency that objects are fragmented into smaller segments, while the sloppy threshold tends to include non-interesting objects, e.g., road marks or asphalts segments (ref. Figure 2.2). The combination of the two thresholds helps to provide fuller (better 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). See Figure 2.3 for an example.

### 2.2.2 New shadow masks

We have used two types of shadow masks; one for removal of tree shadows, and the other for estimation of vehicle shadows. Each of these will be described in detail below.

#### 2.2.2.1 Tree shadows

As mentioned in section 2.1.2 , the previously used method for shadow removal experienced certain problems, and in the current study, we have taken a different approach. Except vehicle shadows, the majority of shadows are located on the edge of the road, and they are caused by trees at the side of the road. A method for shadow removal have been developed, and integrated into the segmentation routine. First, a road edge mask is computed from the manually drawn road mask, using dilation of the road mask with a suitable structuring element. The resulting edge mask is very narrow, so that we may assume there are no vehicles located on the thereby defined road edge. Shadows are removed by simply calculating the overlap with the road edge mask for each dark segment resulting from the main segmentation routine. Segments that overlap the road edge are discarded. This technique helps to reduce the amount of non-interesting segments produced during segmentation, as may be seen for instance in Figure 2.4.

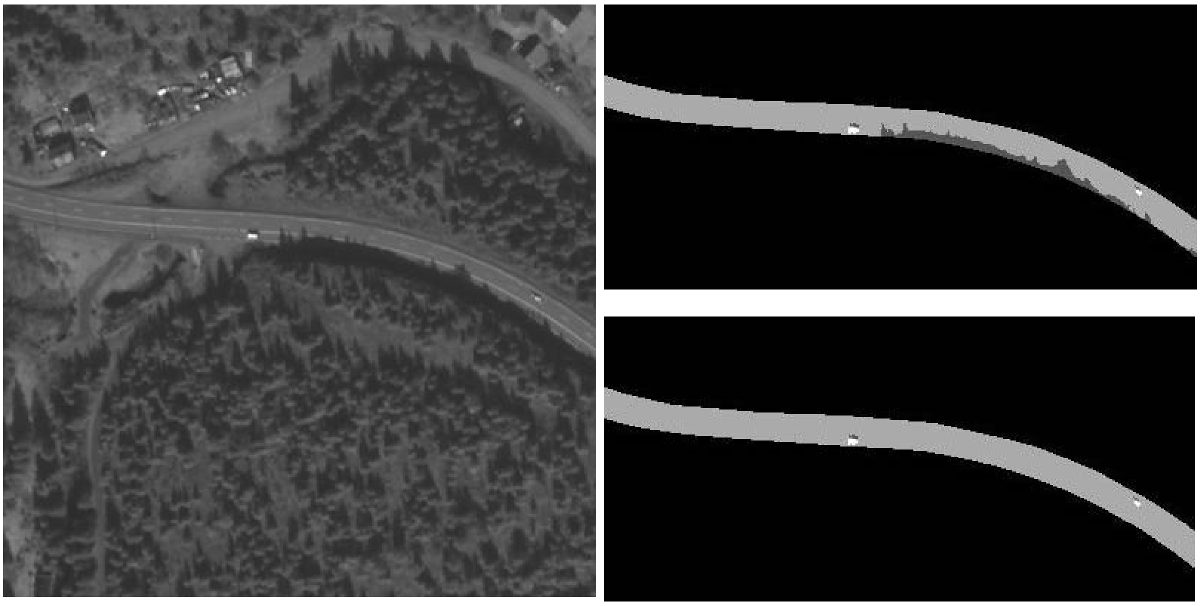


Figure 2.4 The panchromatic image on the left shows a road with tree shadows by the roadside. In the image above on the right, segmentation has been performed without the road edge shadow mask. In the image below, the shadow mask has been used.

#### 2.2.2.2 Vehicle shadows

A new method has been developed regarding vehicle shadows. The *distance to nearest shadow* is estimated as one of the features describing the object, i.e., during the feature extraction routine. Based on information about the sun direction, we first calculate an image mask, indicating the expected shadow zones of the vehicles found in the image.

The expected shadow zone of bright objects is found by dilation[1] of a segmented image, representing bright objects only, with a structure element, representing the expected shadow sector. The direction and the length of the shadow sector are estimated from the sun azimuth and sun elevation, resepectively. These two parameters may vary between different images. We use four different directions to esimate 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

$$\frac{\textit{average vehicle height}}{\tan(\textit{sun elevation angle})}.$$

[1] Dilation is a specific type of mathematical operation performed on an image. Loosely speaking it involves that segments in the image are "dilated", i.e., the segments are grown. A structure element is a small image which describes "the shape of the growing process".

For QuickBird images an average vehicle height of 3 pixels = 1.8 meters was used. Dilation of image $I$ with image (structure element) $S$ is defined as

$$I \oplus S = \{ z \mid \check{S}_z \cap I \neq \emptyset \},$$

where $\check{S}$ denotes reflection of $S$, and $S_z$ denotes translation of $S$ by (a vector) $z$. The dilation result represents an image of the bright objects together with their expected shadow zones, see part b) of Figure 2.5. 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 with a segmented image of dark objects that lie close to a bright object. Whereever 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, and then combining this information with the dark segment image, keeping only the dark segments whose distance to a bright segment is below a preset threshold. Figure 2.5 shows an example.

The result of this process is a shadow mask. This mask is used to estimate the distance to 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.

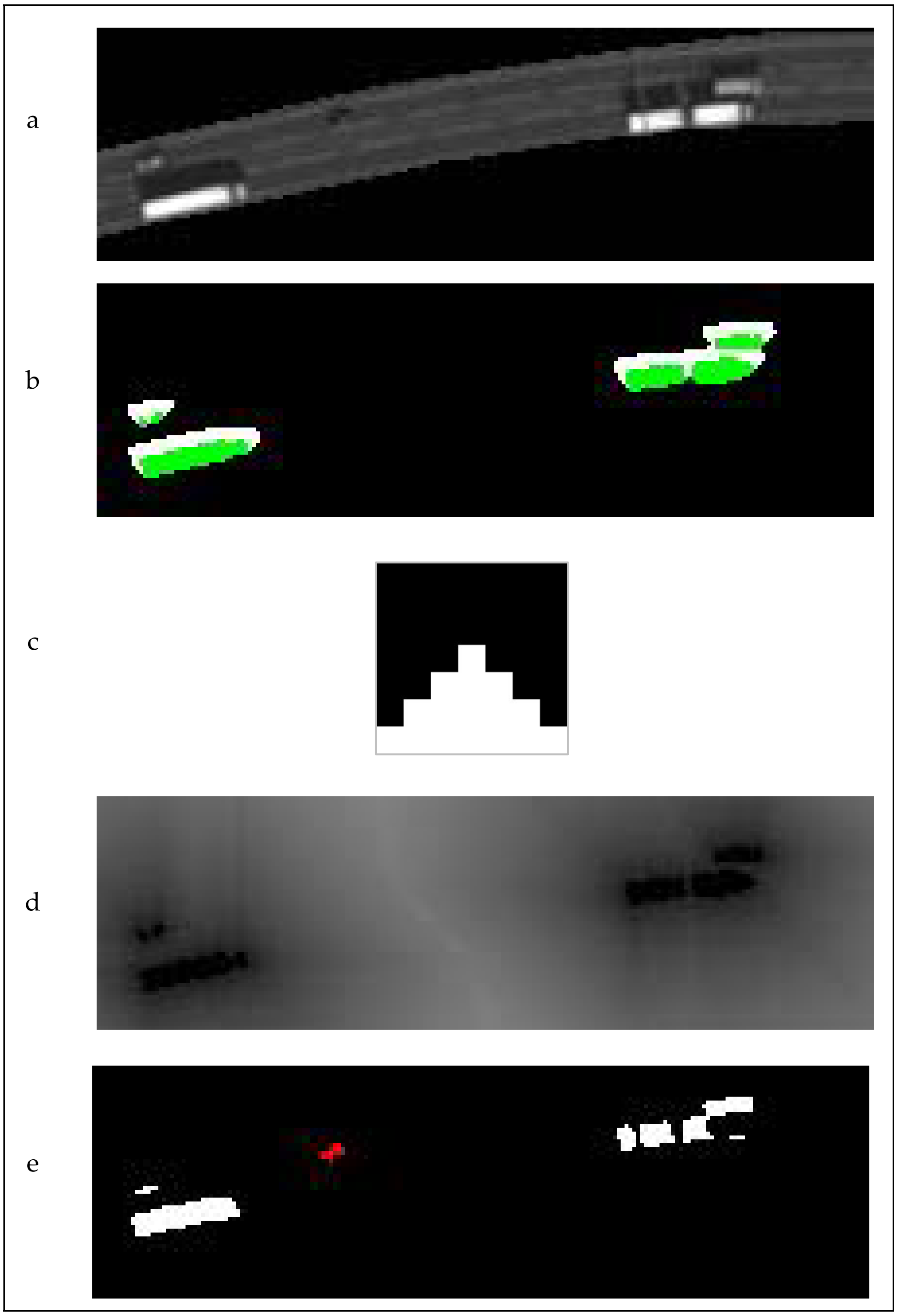



Figure 2.5 The figure illustrates the process of making a shadow mask. a) Panchromatic image of some vehicles and shadows. b) Bright segments (green) and expected shadow zones (white). c) The structure element used for dilation. d) Distance map for bright segments. e) Dark segments – assumed to be shadows (white) and assumed to be vehicles (red).

## 2.3 Classification

The segmentation routine finds most of the vehicle segments, but unfortunately, it also produces other segments that are darker or brighter than the local background, such as road marks, road bridges, or vehicle shadows. In [1] different strategies for vehicle detection were considered, and it was concluded to use a rule-based classifier, followed by statistical classification. In the current work, a direct application of this method - including its choice of features and classification rules - produced (as could be expected) unsatisfactory results. The application of a modified segmentation routine prior to feature extraction may have altered some of the class charateristics. Furthermore, the methods were tested on a larger, more diverse data set. The following is a list of work that has been done:

· The class definition has been extended (the number of object classes has been increased).

· An extensive evaluation of features has been performed.

· The rule-based classifier (pre-classification) has been adjusted.

· The distance to vehicle shadow has been added as a feature.

· Class prior probability distributions have been evaluated.

· The option of adding a reject class has been considered.

The next few sections will describe each point of the list in greater detail.

### 2.3.1 Feature extraction

We have extracted various features describing the shape and the spectral characteristics of the segments. Only the imagination limits the number of features that are possible to extract from the segments of an image. The difficult challenge here is to determine what features best discriminates between the different classes of objects, e.g., what features can tell us the difference between a car and a road mark.

The features we have studied are listed below.

Spatial features:

· length, width, and area of the region’s bounding box

· compactness of the region (the region’s circumference squared divided by the region area)

· elongation

· rectangularity (region area divided by bounding box area)

· spatial spread (calculated using normalized centralmoments)

· ratio between area of region and area of convex hull

· roughness (region circumference divided by convex hull circumference)

· Hu-moments (a set of translation-, scale- and rotation invariant moments)

· angle deviation (difference between direction of the road and first principal direction of the object)

Spectral features:

- mean intensity of the region
- mean image gradient of the region
- standard deviation of the intensity of the region
- mean gradient along the region boundary
- smoothness contrast (local background gradient mean divided by gradient mean of the region interior)

Other:

- distance to nearest vehicle shadow (see section 2.2.2.2)

### 2.3.2 Choice of classifier, class definition and class priors

According to Bayesian decision theory, the probability of error is minimized if we assign a segment 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 minimum-error-rate classification or maximum likelihood classification, and corresponds to the approach that was used in [2] where it was refered to as the Quadratic Discriminant Analysis (QDA). Experiments have been made in order to justify this choice of method. Furthermore, prior probability distributions have been introduced.

The minimum-error-rate classification states that a feature vector $\xi$ should be classified to class $\alpha$ if the posterior probability should be classified to class $\kappa$ if the posterior probability $P(\omega_\kappa | \xi)$ – the probability that the correct class is class $\kappa$, given that the feature vector $\xi$ 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

$$P(\omega_\kappa \mid \xi) = \frac{p(\xi \mid \omega_\kappa)P(\omega_\kappa)}{p(\xi)} = \frac{p(\xi \mid \omega_\kappa)P(\omega_\kappa)}{\sum_{i=1}^{K} p(\xi \mid \omega_i)P(\omega_i)},$$

where $p(\xi | \omega_\kappa)$ is the conditional probability of measuring feature vector $\xi$, given that the vector is a sample from class $\kappa$, $P(\omega_\kappa)$ is the prior probability of class $\kappa$, and $K$ 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 matrix. 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. A common choice is to use equal, constant probabilities for the classes, i.e., the prior probability is $1/K$ for each class, where $K$ is the number of classes. Another common choise is to use the class frequency, i.e., the prior probability $P(\omega_\kappa)$ of belonging to class $\kappa$ is $P(\omega_\kappa) = N_\kappa / N$, where $N_\kappa$ is the number of training samples from class $\kappa$, and N is the number of training samples from all the classes in total. Both choices have been tested, and the latter have shown to provide the best results.



Figure 2.1 Two dimensional scatter plots showing some of the features. (The shadow

In [1] four classes were used; dark noise, bright noise, dark vehicle, and bright vehicle. Through experiments we found that each of these classes consists of a very heteregenous group of objects. A vehicle may be light (small) or heavy (large), and there are many segments that only partially represents a vehicle. Although the remaining segments, i.e., non-vehicle segments, may be characterized as noise in this application, they constitue a wide group of objects, and they obviously do not fit the desired normal distributed class model. Two heteregenous groups of non-vehicle objects are observed on such frequent basis that they have been dedicated a class of their own; heads of bright arrows painted on the road (directions for traffic), and shadows of vehicles. The rest of the objects do not fall into one specific class of shape or intensity. Examples of such objects are

- (parts of) bridges across the road,
- road signs by the roadside or across the road, or shadows of these,
- road marks of various geometric shape (other than the arrow heads),
- (parts of) roundabouts,
- erroneously segmented vehicle objects, grown together in pairs, with road marks, or other objects.

Examples may be seen in Figure 2.7. Typical causes of fragmented vehicles that may be observed in the images of this study include: 1) a bright car is seen as two smaller fragments (probably due to dark interior of a car being visible through the front or rear window), 2) a segment only defines a part of the vehicle because the segmentation is not optimal, usually due to bad contrast, 3) two-colored vehicles are segmented into two segments (most often a bright truck with a dark trailer), 4) other (e.g., split by sign above the road). Examples of different vehicle fragments may be seen in Figure 2.8.

The following classes have been defined.

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

Furthermore, we made experiments where we introduced the possibility of rejecting a segment if none of the class posterior probabilities are 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.

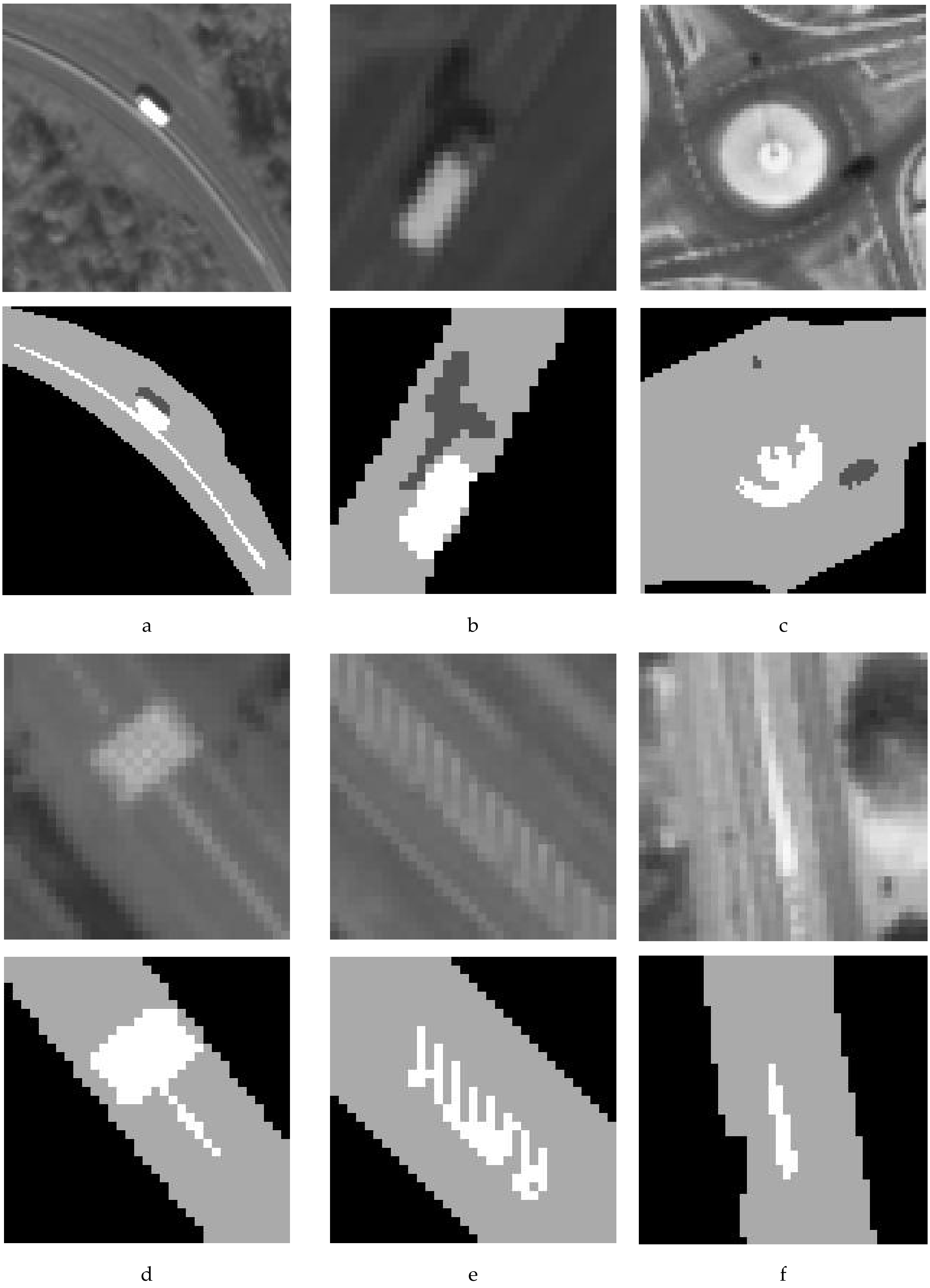
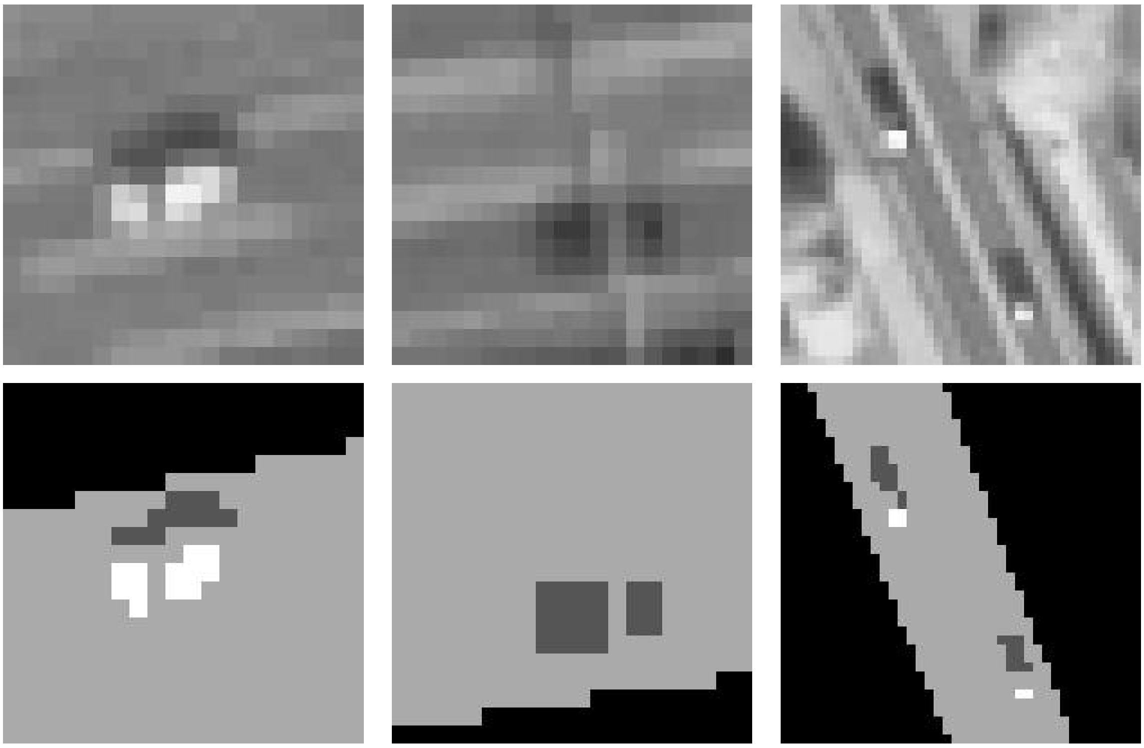




Figure 2.7 Some segments due not fit into any particular shape category or class, and should be rejected. This figure show some examples of image pairs (panchromatic above and segmented below) illustrating this. a) A bright vehicle close to a thick mid line of the road has grown together in a segment. b) A dark car segment has grown together with the shadow of another vehicle. c) Part of a roundabout. d), e) and f)  Untypical road marks.



a b c

Figure 2.8 Vehicles are sometimes fragmented during segmentation. The image pairs illustrate three different reasons for fragmentation. We see pieces of panchromatic images in the upper row with the corresponding pieces from the segmented images below. A) Fragmented bright car due to low light reflection from the front window. B) Dark car fragmented due to signs going across the road. C) Vehicles with two colors, either a bright truck with a dark trailer, or a dark car with sun light reflex from the metallic components or window pane of the car. (Case C is from a different scene than A and B).

### 2.3.3 Feature selection

The PRtools Toolbox in Matlab (PRTools is a Matlab based toolbox for pattern recognition, see [3]) was used for feature evaluation and selection. Different scatterplots was made in order to get an overview of which features discriminate best between the classes, some of which can be seen in Figure 2.9. Ranked individually, the features with most discriminating power are the spectral features, first of all the gradient mean, the intensity mean, the standard deviation of the intensity, and the boundary gradient. The spatial features length, area, spread and Hu-moments, also contribute valuable information. We used a forward floating feature selection method, seeking the optimal number of features, and we used a test set and a training set for evaluation. (Section 3.1 will give a detailed description of the data set). Using the maximum likelihood classification with general Gaussian distributed features for each class, the best results were obtained with the following selection of six features: region intensity mean, region gradient mean, region intensity standard deviation, length, 1st Hu moment, and spatial spread.

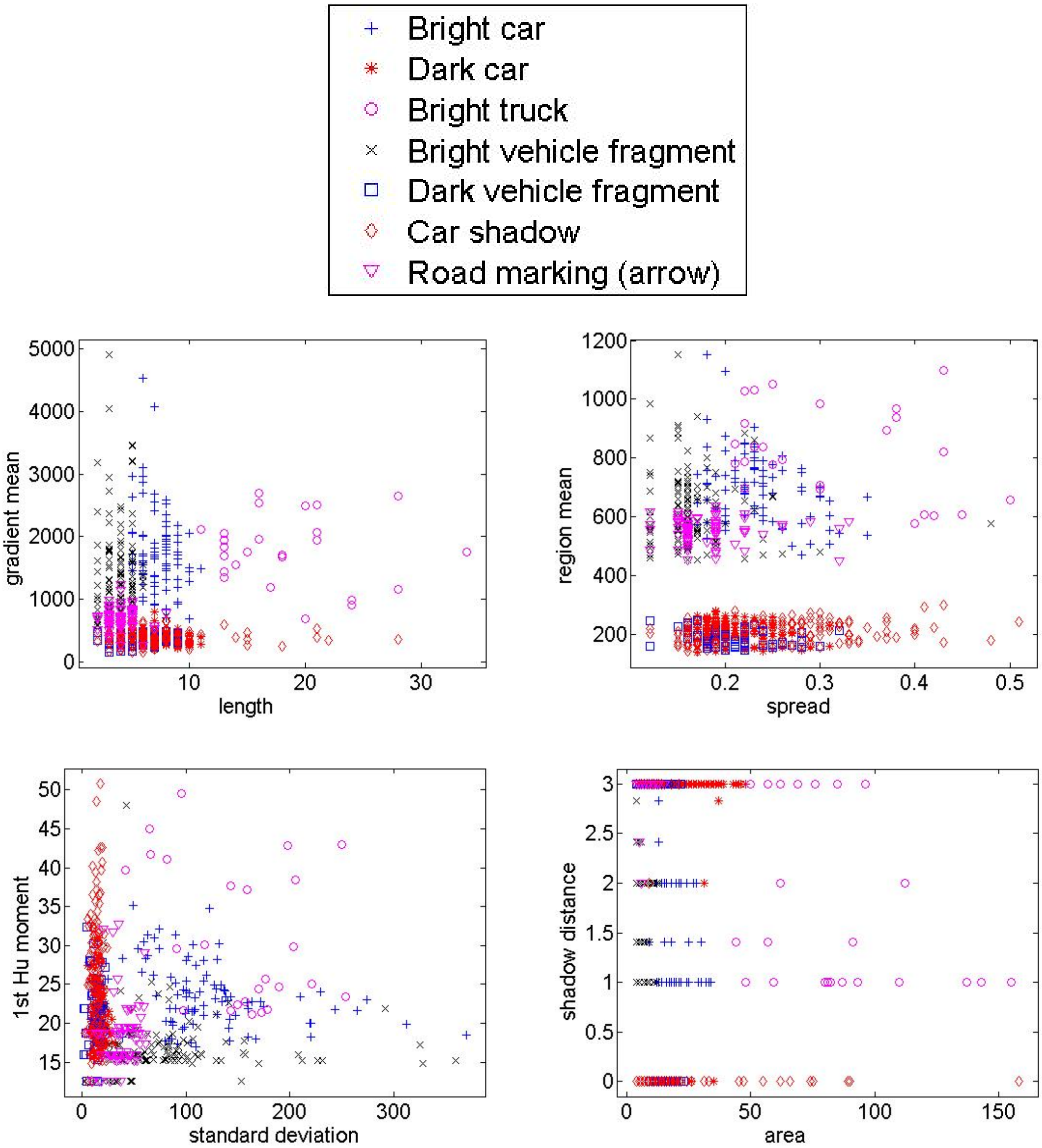



Figure 2.9 Two dimensional scatter plots illustrating the distribution of features. Note that the shadow distance has been truncated at 3.

### 2.3.4 Classification

The classification process is divided into three steps; preclassification, main classification, and postclassification.

#### 2.3.4.1 Preclassification

Before the statistical classification, an initial, rule based classification have been performed in order to discard obvious non-vehicles from the remaining segments. The reason behind making this approach was to reduce the heterogeneousness of the segments. Based on plots of feature vectors with known labels we found upper or lower thresholds for some of the feature values belonging to vehicle objects. These thresholds made the background for a set of simple rules that a segment must fulfill in order to be considered for being a potential vehicle in the next stage. The rule set can be divided in two categories: rules concerning segment size and geometry, and rules concerning intensity and contrast.

The size and aspect ratio of a vehicle segment must lie within reasonable limits. One image pixel corresponds to approximately 0.6x0.6 meters on the ground, and combining this information with experimental results, we have established the following rules:

- A segment is NOT a potential vehicle if its area is greater than 200 or less than four (pixels).
- A potentially large vehicle, i.e., a segment whose length is greater than ten, must have elongation between 1.2 and 4.9.

A test on elongation is not performed for shorter vehicles, as the calculation of length and width may fail to represent the real dimensions of the vehicle, depending on how well the vehicle was defined during segmentation.

The road cover is not a completely homogeneous area when it comes to intensity. Different shades of color on the asphalt, as well as road marks may on some occasions possess an intensity average lying close to what is typical for vehicles. If the region mean is neither particularly bright nor dark, it is most likely not a vehicle. We also expect the gradient of a vehicle segment to be relatively high. The following rules have been applied:

- A segment is NOT a potential vehicle if its mean intensity value is between 300 and 450.
- A potentially dark vehicle, i.e., a segment whose mean intensity is less than 300, must have a minimum gradient mean of 145.
- A potentially bright vehicle, i.e., a segment whose mean intensity is greater than 450, must have a minimum gradient mean of 250.

#### 2.3.4.2 Main classification

The main classification has been applied to all segments characterized as potential vehicles in the preclassification routine. We have used the maximum likelihood classification method as described in section 2.3.2 above. The parameters for the normal distributions were estimated using a training data set, consisting subimages of satellite scenes from Bodø, Kristiansund, Eiker and Sollihøgda (see Section 3.1). With a six-dimensional feature space, there are 27 parameters (6 mean values and 21 (co)variance values) that need to be estimated for each class. For the different classes the number of training samples was: Bright car (123), Dark car (152), Bright truck (37), Bright vehicle fragment (152), Vehicle shadow (206), Road mark – arrow (117).

Once a class description data base (containing the estimated mean and covariance parameters) has been made, the classification routine may be applied to any new image with unknown class labels. The feature vector of each segment is sent into the classification routine, which calculates the posterior probability that this feature vector comes from each of the object classes. If none of the posterior probabilities are above a given threshold (in the current version, this threshold have been set to 0.5), the segment corresponding to the given feature vector is rejected. Otherwise, the segment is classified as the class with the highest posterior probability.

#### 2.3.4.3 Postclassification

The perhaps greatest challenge for the classificator is to distinguish dark cars from vehicle shadows. A shadow cast on the longer side of a vehicle is often hard to separate from a dark vehicle even by manual inspection. The only reason that we recognize it as being a shadow is the fact that it is located in the shadow sector of another vehicle. We therefore sought to reduce the number of vehicle shadows that are wrongly classified as dark vehicles. The postclassification is based on the distance to shadow feature, calculated from the vehicle shadow mask, as described in section 2.2.2.2. Specifically, the class label of a dark car or vehicle fragment is changed to shadow if its shadow distance is zero.

The distance to shadow information was 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 marks do not. The classification of a road mark is changed to bright vehicle fragment if its distance to shadow is less than three pixel units.

# 3 Image data and road study sites

## 3.1 Description of the data set

The objective is to see if detection of vehicles from satellite images could be used to estimate the amount of traffic on certain roads. To be able to detect vehicles, satellite images with high resolution are required. We have chosen the QuickBird satellite with minimum resolution of 0.6 m. By searching in the QuickBird image archive we have found a number of images covering different parts of Norway for a period between 2002 and 2006. The images are taken in the summer season without snow on the roads and enough sunlight to be able to detect vehicles. From quicklooks we have found the main roads, or actually the parts of main roads, which are covered by these images.

From a list of these roads, the Norwegian Public Roads Administration has studied the traffic counting data to see if there is a match between the image acquisition dates and the dates of traffic counting for the corresponding roads.

Six different roads were selected and seven subsets of satellite images were ordered. The selected sections of QuickBird images are listed in Table 3.1 with the name of the roads to be tested, and the date and time of the acquisition.

| Location | Roads | Date / Time UTC | Upper left latitude | Mean sun elevation angle | Mean off nadir view angle |
|---|---|---|---|---|---|
| Sennalandet | EV6 | 03.05.2006 10:35 | 70 | 36 | 9,9 |
| Bodø | RV80 | 21.07.2003 10:32 | 67 | 43 | 4,4 |
| Kristiansund | RV70 | 19.06.2004 10:56 | 63 | 50 | 7,9 |
| Østerdalen north | RV3 | 10.08.2004 10:39 | 62 | 43 | 7,3 |
| Østerdalen south | RV3 | 10.08.2004 10:39 | 62 | 43 | 5,3 |
| Eiker | EV16 | 07.06.2002 10:42 | 60 | 53 | 12,9 |
| Sollihøgda | EV134, RV35 | 10.05.2002 10:32 | 60 | 47 | 12,5 |

Table 3.1 Selected subset of QuickBird images.

There are two image sections covering RV 3. The counting station is located in the northern part, Østerdalen north. The part of RV 3 in the southern image, cannot be used to confirm traffic counting, but can be used for training purposes. EV 134 in the Eiker image did not have traffic counting at the acquisition day, but there are many vehicles of different types on this road, and they can be used for detection training. RV 3 runs mainly in direction from south to north, and EV 134 mainly from east to west. Then there are a number of examples of vehicles seen from

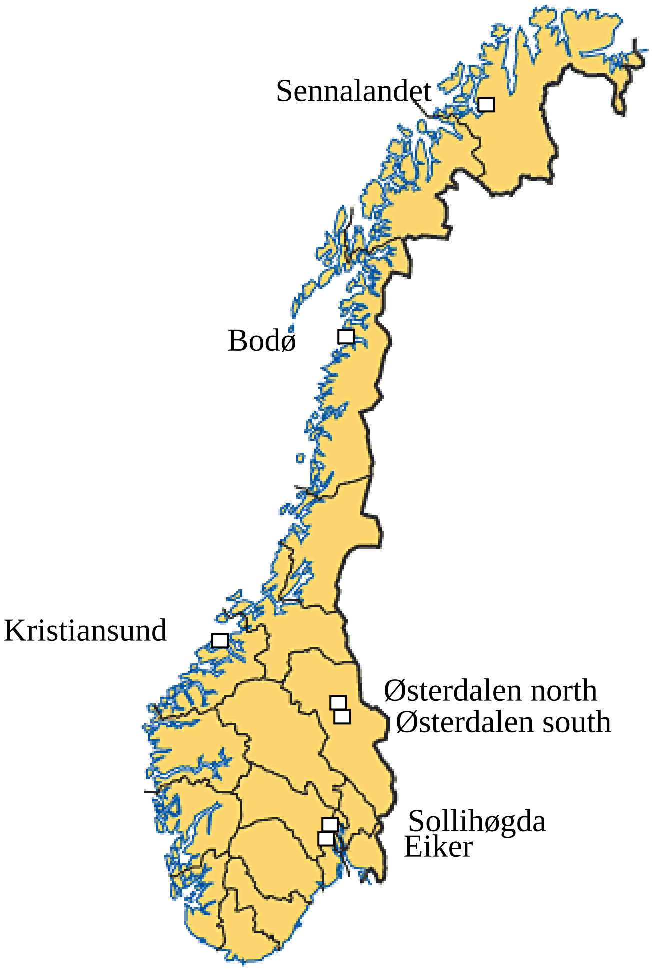



Figure 3.1 Location of QuickBird images

different directions to be used in detection training. For EV 134 the viewing direction from the satellite and the large viewing angle, give a lot of examples of the effects of vehicle shadows.

The location of the used satellite images are marked on the map of Norway in Figure 3.1. The selected subsets of QuickBird images with the roads marked are shown in Figure 3.2.

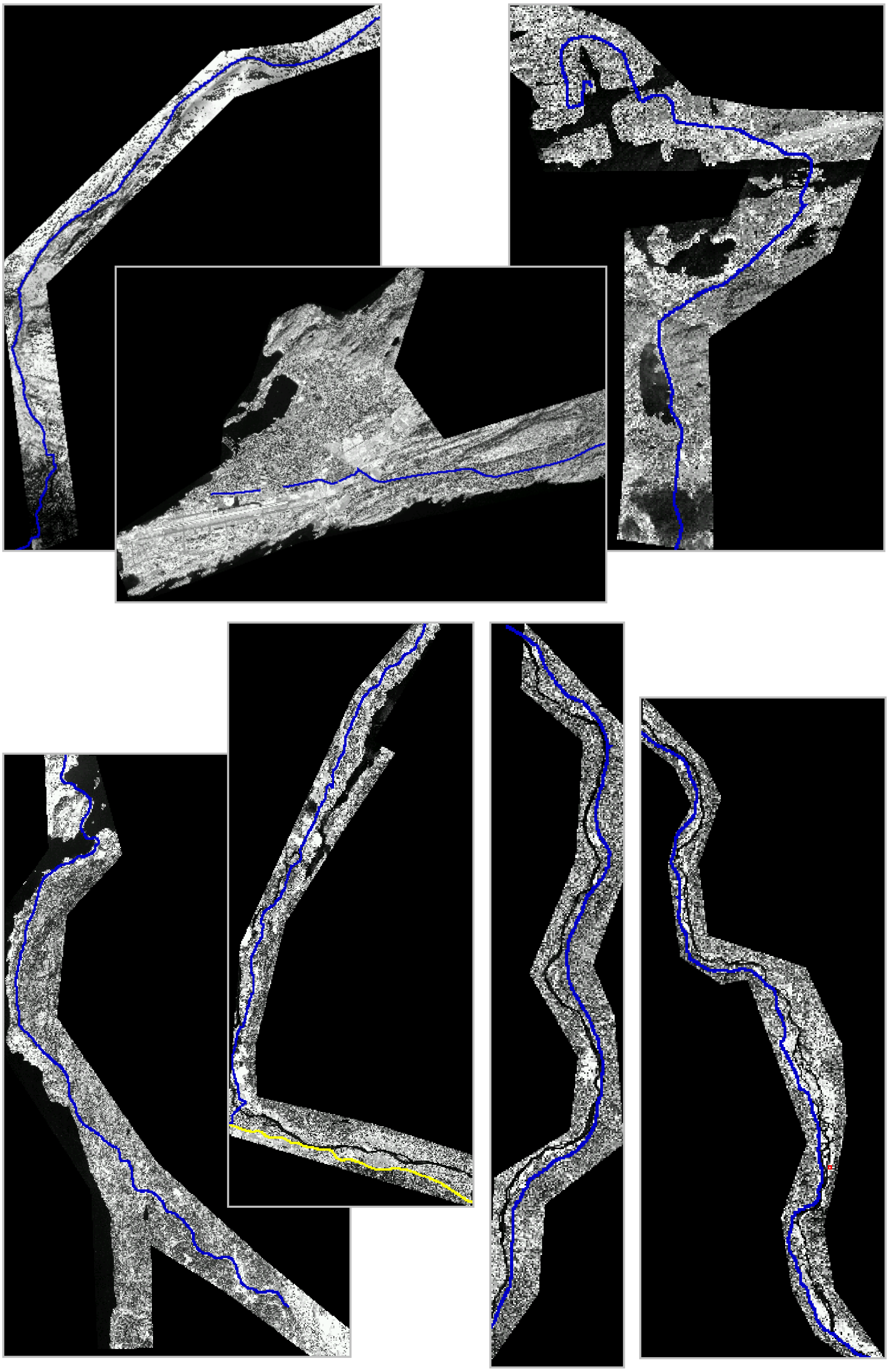


Figure 3.2  Subsets of QuickBird images with marked roads.

## 3.2 Making road masks

The idea of making road masks is related to the problem of automatically detecting roads in a satellite image. The problem of finding vehicles in the image will be substantially reduced if you know the position of the roads. A road mask is a subset of an image, consisting ideally of only the pixels which cover the road. However, in the real world it will also contain some pixels outside the road borders. This is necessary to include tall vehicles which will be imaged outside the road area, when looked upon with a large viewing angle from the satellite. In Figure 3.3 one can see tall vehicles being imaged outside the road borders. The left vehicle is covering the outer road marking line. The right vehicle is partly covering the marking line, and the shadow is stretching outside the road. The shadow of the vehicles can be used in the classification process, and the complete shadows should be included inside the road mask.

Using ENVI and IDL there are a couple of different ways to create road masks. These will described below.

Figure 3.3 Tall vehicles are imaged outside the outer road markings.

### 3.2.1 Manual digitalization

The most direct way to make a road mask is to perform a manual digitalization. In ENVI one can make one or more ROIs (region of interest) describing the road area, by drawing lines that describe a polygon with the cursor in a viewer. The ROIs can be transformed to a class image interactively or with a call in an IDL program. The pixels within an ROI will have the same value, and the ROIs will have different values. To make a road mask, all class values are converted to 1. The pixels outside the road will have value zero. The new raster image is multiplied by the original image to get the final road mask. The multiplication can be made in a program or interactively by use of the command Band Math.

The advantage of manual digitalization is that you can follow the road exactly as it is presented in the image. You can exclude or include bus stops and other extensions as you like and you have full control on roundabouts, bridges and tunnels. Tunnels and bridges crossing the roads should be excluded to reduce the problems of vehicle detection. You can draw the lines in such a way that all the vehicles you see on the road are inside the ROIs.

When you digitalize the original image, you do not have to make an extra resampling of the image to get correct coordinates. A resampling could make small changes in the information of the vehicles on the road and reduce the difference between the road and vehicles and making them more difficult to detect.

However, if you need to take into account information of position of special locations on the road, like county borders, counting sites etc., you may need to transform the geographical coordinates to the image coordinates.

One disadvantage with manual digitalization is that it is a tedious job, which needs concentration. If you have many long roads to digitize, you may wish after a while that it was possible to do it automatically.

### 3.2.2 Automatic generation of road masks

#### 3.2.2.1 Input data

The Norwegian Public Roads Administration keeps records of the main roads as vectors describing the midlines with additional information about the properties of the road, stored as shapefiles. The vectors are drawn in UTM projection zone 33, WGS-84 for all Norway. Among the registered properties are the widths of the carriageway, the road surface, and the total road. One way to make a road mask is to use the vector as a source for growing a raster buffer around it, with the width given in the parameters. The Norwegian main roads are divided into parts called main parcels. In the parameter list the road is divided into smaller parts, with start and end point given in distance (in meters) from the start of the main parcel. For each of these smaller parts, the road width is given in meters. The length of the parts may vary, but many parts have a length of a multiple of 500 meters. This does not mean that the road changes width each 500 meter. The width has been measured at the road marks which are placed with 500 meter intervals. How the width varies between these markings can not be found from the vector data.

#### 3.2.2.2 Making a buffer mask

When creating a buffer around the road vector, one would have to find the positions along the vector for each change of width. The image coordinates must be calculated from the length parameter. The width data cannot be used to make a road mask which exactly matches the road part by part. A better way to do it is to find the maximum width of the road and make a buffer with this width (+ a couple of meters) for the complete road length.

ENVI has an interactive command which creates a buffer around a vector line. Open the vector in a vector window and select File->Calculate Buffer Zone. You have to select an associated data (raster) file. You can also select Spatial Subset if you want to treat only a part of the road. Then you set the Maximum Distance (maxdist) in pixels. The pixels in the created buffer image have values equal to the distance from the input vector. Pixels with a distance equal to or larger than maxdist, will all have value maxdist. This means that for a specific width in meters you have to find half this width in pixels.

From the buffer image one can make a buffer mask by changing the value of all pixels with value maxdist to 0 and all others to 1. This can be done with a program or interactively by the

Figure 3.4  Road center lines has gaps at roundabouts.

Band Math command. A complete road mask is created by multiplying the original image with the buffer mask. To do so the two images have to be in the same projection.

There are some problems with automatic creation of a buffer mask.

- The road vectors usually have gaps at roundabouts, see Figure 3.4.
- There may be gaps at other places where there should not be gaps.
- The road vector is drawn also where the road is invisible from the satellite, like in tunnels and under crossing bridges, see Figure 3.5.

After a buffer mask has been made, manual editing will be necessary. This editing cannot take place before the buffer has been transformed to the projection of the satellite image (or vice versa).

#### 3.2.2.3 Rectification

The satellite images are delivered by the producer in an approximate UTM zone 33 projection, datum WGS-84. The images have to be adjusted to match the road vectors in the shapefiles. To avoid resampling of the images, the road buffer mask can be rectified to match the original image. In ENVI this can be done by selecting

Map->Registration in the main menu. The original image is selected as the 'Base image' and the buffer mask as the 'Warp image'. A number of ground control points (GCPs) are chosen in the original image. For each point the image coordinates are saved. To find the position of the corresponding point in the buffer mask image, one has to use a map. Digital maps of Norway can be found at the website Norgesglasset produced by Statens Kartverk. The coordinates

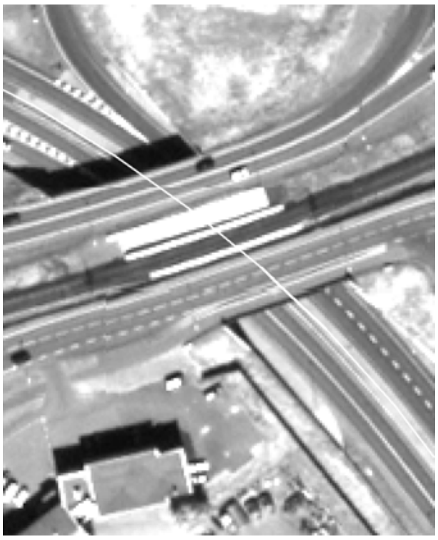


Figure 3.5 Road center line is drawn upon a crossing bridge.

found for the corresponding point in the map must be transformed to image coordinates. This can be done by moving the cursor to the position with the actual coordinates in the buffer mask image and read the image coordinates. These coordinates are stored in the GCP-list together with the coordinates from the original image. When you have a number of GCPs, the buffer mask image can be warped to match the original image. Depending on the number of GCPs you can have different types of transformation. The number of points needed to get a satisfactory transformation, depends on the terrain and on the viewing angle from the satellite.

This image-to-image matching does not involve knowledge about the terrain and satellite parameters. ENVI also has an Orthorectification command which takes these parameters into consideration. Among the files delivered together with the panchromatic and multiband images, there is a file with extension .rpb, which contains the necessary information about the satellite camera and its position. This file is input to the orthorectification process together with a digital elevation model. In ENVI the panchromatic image can be rectified to match the map, but not vice versa. When using this orthorectification process, the satellite image will be resampled. This may make it more difficult to detect the cars, but that is not for sure and has to be closer studied.

### 3.2.3 Results

For the road maps from Bodø, Kristiansund, and Sollihøgda, road masks have been made by manual digitalization. Each road has been digitized by dividing it into shorter parts when drawing the ROIs. Typical division points have been made at tunnel entrances, bridges crossing the road, and other large crossings. The borders have been drawn outside the side markings of the roads where they are visible. At places where trees or buildings are covering the view of the road, the position of the border of the road has been estimated. At places where the image of a car extends outside the road, the car has been included in the road mask.

For the other images it has been tried to make road masks from automatic buffering. A test was made for Bodø, where there already existed a manual digitized road mask. A buffer with maximum distance of 10 pixels was made. With a pixel size of 0.6 x 0.6 meters, this corresponds to a total road width of 12 m. With 5 control points the buffer mask was matched with the image. Controls showed that the mask matched the road satisfactory. The road is situated in a flat area, and the mean viewing angle of the image is 4.4°.

For Østerdalen north the viewing angle is 7.3°. The road is located in the bottom of a valley, running mainly in one direction, from north to south. The elevation of the road is steadily decreasing from north to south, 100 m over a distance of about 30 km. 6 control points close to the road were selected. These were enough to rectify a road mask which matched very well with the image.

Figure 3.6 Center line for four lane road.

For Østerdalen south the mean viewing angle is 5.3°, the terrain is quite similar to Østerdalen north. 4 control points were enough to create a satisfactory road mask.

For Sennalandet the mean viewing angle is 9.9°, the road is starting near the ocean in south and then mounts to about 380 m. The track goes first north and then east. Here it was difficult to find control points to get the buffer mask matching the complete road. 11 control points were necessary before a usable transformation was made. And even then the matching was not perfect.

In the Eiker image the mean viewing angle is as large as 12.9°.The image was divided into two parts, one containing EV 134 and another including RV 35. This was done to reduce the data in each image and to simplify the rectification. Even with 11 GCPs, the buffer mask for EV134 did not match the image satisfactory. An orthorectification of the total image including a digital elevation model, but without using control points, gave an almost satisfactory result. The rectified image matched well with the initial buffer mask. Using this image instead of the original, the problem of resampling occurs, but this is perhaps not a significant problem after all. It needs to be tested.

Along EV 134 another problem was revealed. Over a certain distance the road changed from two to four lanes. The vector was drawn in the middle of the two westbound lanes and not in the middle of all four, see Figure 3.6. The total width is greater than the maximum width given in the shapefile.

### 3.2.4 Discussion

From the tests it seems as if the geo correction works well with low values of the viewing angle and gets more difficult as the angle increases.

The process of rectification creates a lot of manual work. The control points have to be carefully selected, and the coordinates have to be found on a map with satisfying accuracy. In the Eiker case, the existing high-scale maps were not all recently updated and lacked details which could have been used as control points. The road EV 134 was not even drawn on the high-scale map over a certain distance, but existed on maps in lower scale.

The maps on the Internet are in other projections than the shapefiles, and an extra coordinate transformation is necessary.

During the tests a couple of errors in the ENVI programs were found. The storing of control point coordinates was not performed properly, and there could be errors in the positions of the control points of as much as 5 pixels. This may have influenced the rectification process.

An automatic production of road buffer masks will create masks that have to be edited. Tunnels and crossing bridges should be removed. All gaps have to be closed, roundabouts have to be included. There may be necessary to increase the width due to incorrect width data and errors in the positions of the road.

It is not easy to get a correct rectification of the buffer mask, especially if the satellite view angle is large and the roads passes through terrain with large changes in elevation. There will probably be errors, small or large, and it will be necessary to make manual changes to have the buffer mask covering the road completely. To reduce this problem a wide buffer should be made. Then one will get a mask which covers the road, but which also covers a zone outside the road on one or both sides. This may introduce a lot of false classifications of vehicles.

If there are just a few roads to be treated in an image, it will probably be most effective to do a manual digitalization. You do not have to perform the rectification, and you avoid all extra editing, which usually will cost a lot of time.

# 4 Validation

It is not straightforward to assess the results of the automatic vehicle detection algorithm. We have used three different approaches in order to obtain a validation of the suggested methods; comparison with manual counts, comparison with data from in-road equipment, and evaluation of the classification error. Each of these approaches will be explained in detail below. Results are also provided.

We have been processing subsets of the satellite images that were described in Section 3.1. Table 4.1 gives an overview of what images have been used for algorithm development, training the classifier, testing the classifier, and other validation activities, respectively. The Østerdalen south image was used for road mask generation (section 3.2), but as there are no in-road stations here, the image has not been studied further. The Sennalandet image has a somewhat different type of intensity distribution than the other images. The scene contains snow, which reflects much more light than a bare ground, and the intensity values are higher than in the other images in our data set. Thus, the segmentation routine behaves a bit differently. With only slight modifications of the segmentation, the automatic method was able to detect some vehicles in this image. However, we only report manual counts from this image for now. See section 5.3 for further discussion of this case.

| Location | Image subsets are used for: | | | |
|---|---|---|---|---|
| | algorithm develop-ment | statistical training | statistical testing | other validation activities |
| Sennalandet | | | | x |
| Bodø | x | x | | |
| Kristiansund | x | x | x | x |
| Østerdalen north | | | x | x |
| Eiker | | x | x | x |
| Sollihøgda | x | x | x | x |

Table 4.1 Overview of image data usage.

## 4.1 Comparison with manual counts

The main objective of the automatic vehicle detection method is to count the number of vehicles that are present in the satellite image. It is of course possible to count the number of vehicles by manual inspection of the image. However, the result of a manual count is not necessarily correct, as it depends on the person who performs the count, and how he/she perceives the image. Even with a resolution of 0.6 meters, some vehicles are hard to detect due to low contrast, shadows, etc. Furthermore, sometimes there are objects in the image that may resemble a vehicle, e.g., a road mark or a patch of new asphalt, in which case two different persons may reach a different number of vehicles when counting. Thus, we let two individual persons perform manual counts. The numbers were then compared to each other and to the number found by the automatic algorithm. Results are given in Table 4.2 and Table 4.3. The location names refer to the satellite image from which the subimage was taken, see Table 3.1.

Two of the images (Kristiansund and Sollihøgda) had in-road counting stations at two different locations.

| Location | Length of road segment (m) | Manual vehicle count | | |
|---|---|---|---|---|
| | | Person 1 | Person 2 | Consensus |
| Sennalandet | 19718 | 13 | 16 | 12 |
| Kristiansund # 1 | 1055 | 22 | 22 | 21 |
| Kristiansund # 2 | 5775 | 32 | 33 | 32 |
| Østerdalen north | 31779 | 44 | 41 | 41 |
| Eiker | 7836 | 57 | 55 | 54 |
| Sollihøgda # 1 | 7819 | 63 | 63 | 62 |

Table 4.2 Consensus between two different manual counts. By consensus we mean that the same vehicle has been counted by both persons.

| Location | Objects classified as vehicles | Correctly classified as vehicles | | Correct classification rates | |
|---|---|---|---|---|---|
| | | Person 1 | Person 2 | Person 1 | Person 2 |
| Kristiansund # 1 | 17 | 14 | 14 | 82,4 % | 82,4 % |
| Kristiansund # 2 | 22 | 21 | 21 | 95,5 % | 95,5 % |
| Østerdalen north | 80 | 32 | 32 | 40,0 % | 40,0 % |
| Eiker | 39 | 35 | 32 | 89,7 % | 82,1 % |
| Sollihøgda # 1 | 64 | 48 | 47 | 75,0 % | 73,4 % |
| Sollihøgda # 2 | 26 | 24 | 24 | 92,3 % | 92,3 % |

Table 4.3 Automatic vs. manual vehicle count.

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 4.3

the column “Objects classified as vehicles” presents the number of vehicles found by the algorithm counting fragments that belong together only once.

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 marks and tree shadows. In the Østerdalen image, there are 29 tree shadows and eight road marks that are wrongly classified as vehicles. In the Sollihøgda # 1 image, the corresponding numbers of tree shadows and road marks are three and twelve.

## 4.2 Comparison with in-road counts

As mentioned in the introductory part of this note, today’s method of generating traffic statistics relies on point measurements from in-road stations. The Norwegian Public Roads Administrations has provided data from point measurements at locations within the areas covered by the satellite image data (see section 3.1 for a description of the study sites). The in-road equipment supplies information about how many vehicles that passed the counting station during a time interval of one hour, for each whole hour throughout the day. The average vehicle speed is also provided. These numbers were used in order to estimate how many vehicles that are expected to be present in an image that is acquired during the same time as the measurements took plave.

For each point measurement, a stretch of road on each side of the station was selected, and the corresponding image subset was extracted from the original satellite scene. Automatic and manual vehicle counts were performed on these subimages and compared to the number to the in-road counts. Table 4.4 presents number of vehicles in a snap-shot situation. More specifically, the table contains 1) the number of vehicles that was manually counted in the image, 2) the number of vehicles that is expected to be present in the image based on in-road counts, and 3) the number of vehicles that was found in the image by the automatic algorithm. We use both in-road counts from the time interval in which the (snap-shot) image acquisition took place as well

| Location | Length of road segment (m) | Time of image acquisition | Manual count in image (person 1) | Predicted number of vehicles in image 10-11 UTC | Predicted number of vehicles in image 11-12 UTC | Objects classified as vehicles |
|---|---|---|---|---|---|---|
| Sennalandet | 19718 | 10:35 | 12 | 10 | 9 | - |
| Kristiansund # 1 | 1055 | 10:56 | 22 | 25 | 25 | 17 |
| Kristiansund # 2 | 5775 | 10:56 | 32 | 27 | 28 | 22 |
| Østerdalen north | 31779 | 10:39 | 44 | 51 | 40 | 80 |
| Eiker | 7836 | 10:42 | 57 | 57 | 67 | 39 |
| Sollihøgda # 1 | 7819 | 10:32 | 63 | 58 | 61 | 64 |
| Sollihøgda # 2 | 6139 | 10:32 | 30 | 38 | 41 | 26 |

Table 4.4 Comparison with in-road counts.

| Location | 10 - 11 UTC | | 11 - 12 UTC | |
|---|---|---|---|---|
| | In-road count | Predicted based on manual count | In-road count | Predicted based on manual count |
| Sennalandet | 49 | 57 | 42 | 59 |
| Kristiansund # 1 | 1228 | 1089 | 1246 | 1087 |
| Kristiansund # 2 | 318 | 381 | 330 | 379 |
| Østerdalen north | 140 | 123 | 111 | 123 |
| Eiker | 491 | 491 | 578 | 491 |
| Sollihøgda # 1 | 504 | 548 | 536 | 554 |
| Sollihøgda # 2 | 533 | 421 | 576 | 422 |

Table 4.5 In-road counts vs. predicted number of vehicles based on manual counts.

as the following time interval. Table 4.5 presents number of vehicles that pass the in-road measurement point during a one hour period. Again, we use the one hour interval that contains the time of when the image was acquired in addition to the following one hour interval. The table also presents the number of vehicles that would pass the in-road station during one hour if the manual count is representative for the whole time period. We estimate this number based on the number of cars that was counted manually in the image, the length of the road segment, and the speed of the vehicles.

All of the images were acquired in the time interval between 10 and 11 UTC. However, some of the images were acquired late in this interval, and it is reasonable to compare them also to counts in the following time interval, i.e., to counts made between 11 and 12 UTC. This is because the traffic may be increasing or decreasing over time, and this is a continuously ongoing process.

## 4.3 Classification performance

The performance of the classification algorithm has also been reported. If a vehicle is missed in the segmentation step, it will not be classified. Thus, the evaluation of the classification

| 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 4.6 Classification confusion matrix.

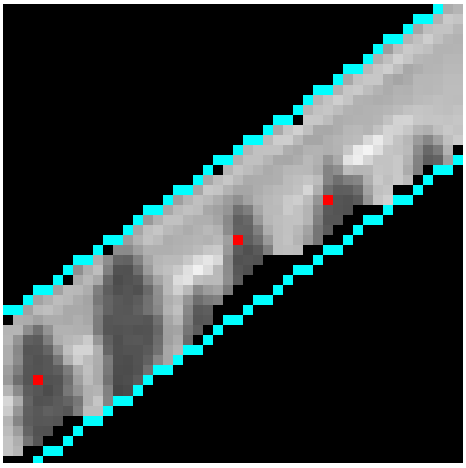
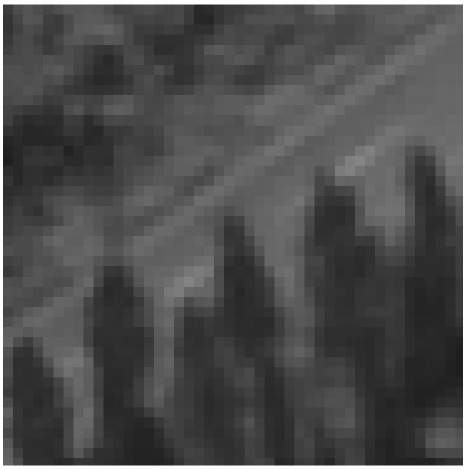
 

Figure 4.1 In the left image we see that tree shadows are wrongly classified as vehicles (red points). This is the masked panchromatic image, i.e., the background , which is defined by the road- and vegetation masks, is black. The road edge is marked in cyan.  The corresponding panchromatic image is on the right.

performance alone does not tell us how many vehicles are missed by the algorithm in total. Nevertheless, the classification method is an important factor in the vehicle detection algorithm, and the overall performance of the algorithm relies on a good classification result.

Table 4.6 is a so-called confusion matrix for the classification result. The reported results are from processing of a selected set of subimages from the Kristiansund, Østerdalen, Sollihøgda, and Eiker images (see Table 4.1). The bright vehicle classes have been merged. The given labels refer to the labels that were assigned to the segments that were found and classified by the automatic algorithm. An operator was then manually led through the segments and asked to edit those labels where the program was wrong. The true labels refer to the manually edited labels.

We see that most of the vehicles (89.6%) that are found in the segmentation routine are correctly classified by the algorithm. The segments that should have been rejected seem to contribute the most to the classification error. Confusion between dark vehicle and vehicle shadow, or bright vehicle and road mark, is also important. The total classification rate was 70.6% including reject segments, and 88.7% *not* including reject segments. If we only consider two classes; vehicle or non-vehicle, the classification rate was 81.0%.

## 4.4 Discussion of the validation results

The results in section 4.1 indicate that 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, see Figure 4.1. 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.

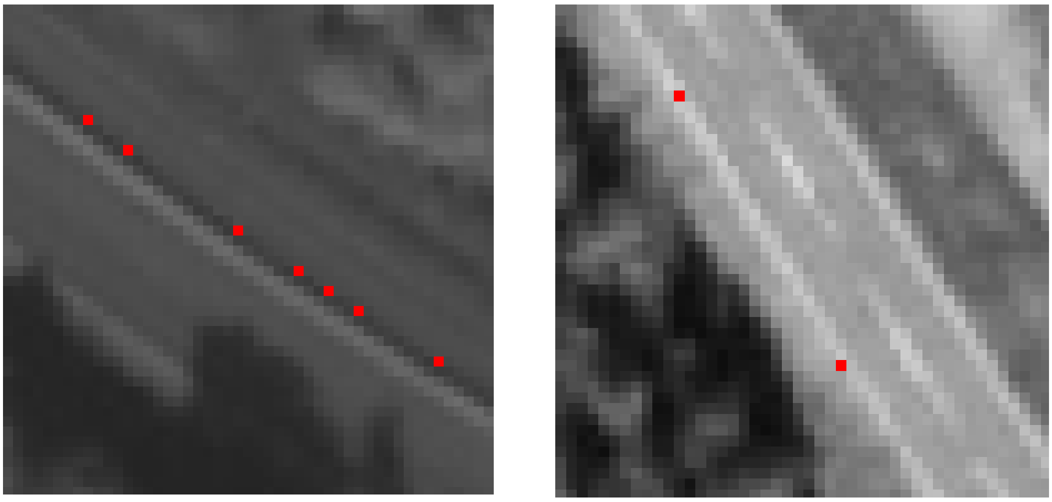


Figure 4.2 The automatic method sometimes overestimate the number of vehicles. The red dots mark segments that have been classified as vehicles. Left:: Dark road marks, or shadows of a in-road fence? Right: Bright road marks.

The validation shows that the automatic algorithm finds many segments that do not belong to any of the defined classes. Examples of such segments were illustrated in Figure 2.7. Road marks also come in many shapes and with different gray tone intensities. The type of road marks that are most frequently found during segmentation is the arrow shaped road marks. Mid-line road marks often has a slightly lower contrast and are therefore avoided in the segmentation. However, this is not always the case. As noted in section 4.1, the Sollihøgda # 1 and the Østerdalen images also contain many segments that are wrongly classified as vehicles. These problems are illustrated in Figure 4.2.

In four out of seven images the automatic vehicle detection underestimated the number of vehicles. This is mainly caused by the fact that some vehicles are missed by the segmentation. In the Eiker image, many vehicles are probably missed due to the cloud shadow covering a large part of the image, see Figure 4.3 and Figure 4.4. The manual vehicle counts were performed using the ENVI software. The built-in contrast enhancement filters were used in those parts of the images where it was needed. This may explain why some of the vehicles that were found by both manual inspectors are *not* detected by the automatic algorithm. It may suggest that the segmentation algorithm lacks robustness and has problems to adapt to different lighting conditions.

Figure 4.3 A part of the Eiker image is shown on the left. Note the cloud and cloud shadow in the image. On the right we see a stretch of road from the left image where the lighting conditions are good and the image is clear. Compare with Figure 4.4.

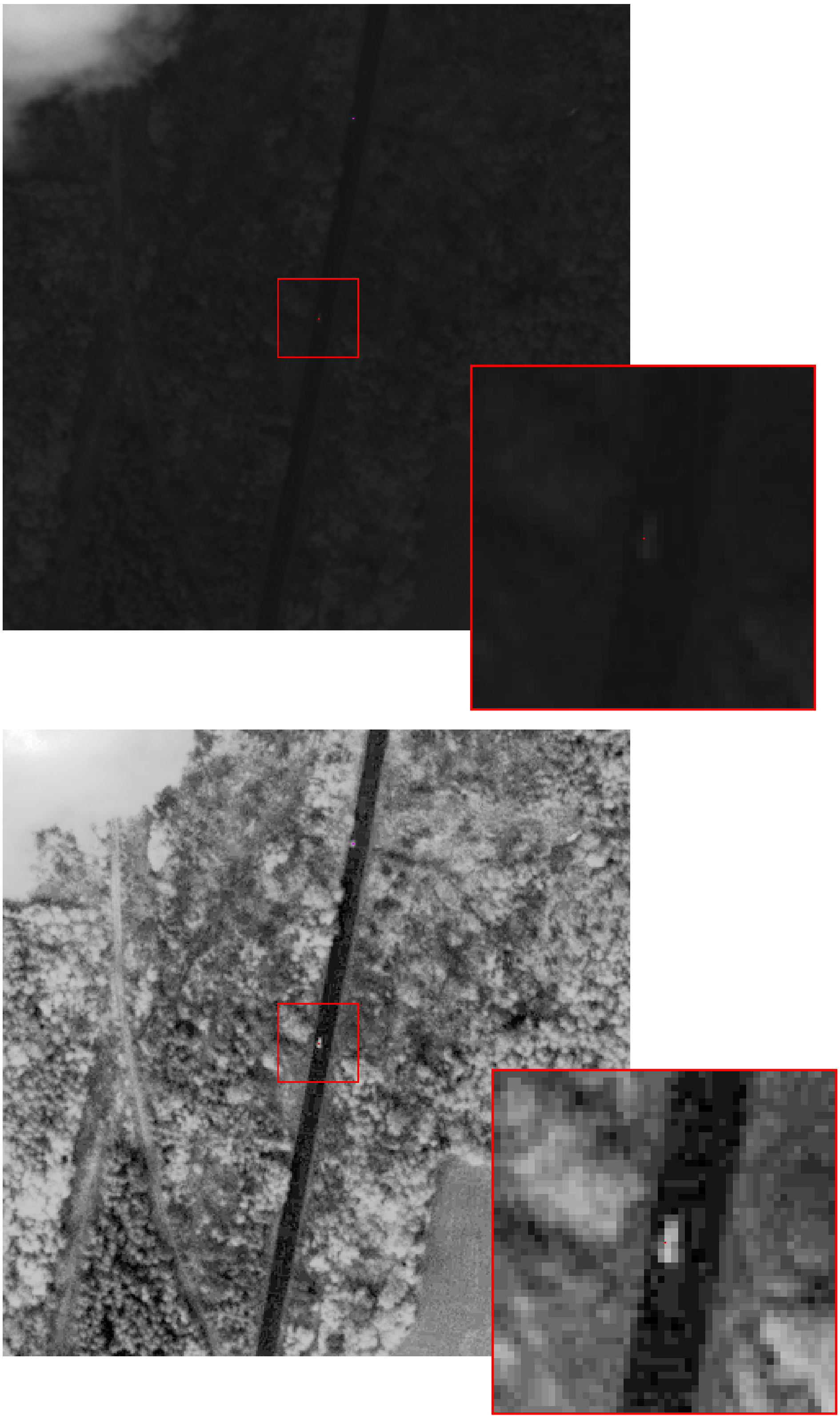


Figure 4.4 The images show another subsection of the Eiker image from Figure 4.3. This subsection illustrates a stretch of road that is covered by the shadow of the cloud. Above we see the image as it is (no contrast enhancement). Below we see the same image after that Gaussian contrast enhancement has been applied to it.

# 5 Conclusions

## 5.1 Summary

In this report we have presented an approach for detection of vehicles from high-resolution satellite images. The methodology has been developed using a selected set of images of Norwegian roads, but we have attempted to make it as general as possible. The presented algorithm was developed as a part of the "SatTrafikk" project, and is based on the approach that was suggested in the ESA project "Road Traffic Snapshot" from 2006-2007. The automatic vehicle detection method consists of a segmentation step, followed by feature extraction and classification of the segments.

The segmentation works on the panchromatic image after masking, i.e., everything except the road is considered to be image background. Image segments representing potential vehicles are found 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. We also use one context based feature; the distance to nearest vehicle shadow. 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).

The total classification rate was 70.6% including reject segments, and 88.7% *not* including reject segments. If we only consider two classes; vehicle or non-vehicle, the classification rate was 81.0%. According to manual vehicle counts, 70.2% (person 1) or 68.5% (person2) of the segments that were classified as vehicles were correctly classified. In the images Kristiansund #1 and #2, Sollihøgda # 2, and Eiker, the number of vehicles was underestimated. Compared with manual counts, the automatic method found respectively 77%, 69%, 87%, and 68% as many vehicles as person 1. In the images Østerdalen north and Sollihøgda #2 the number was overestimated, 157% and 110%, respectively. A very crude explanation is:

· Overestimation is caused by a heteregeneous group of segments that do not belong to any class, but the method does not know how to reject these segments.

· Underestimation is caused by abnormal lighting conditions in (parts of) the image or vehicles that for other reasons have low contrast to the surroundings.

Some ideas of how these problems may be solved will be discussed below (section 5.3).

## 5.2 Main algorithm improvements

Many dark segments on the road are shadows of bright vehicles. We have proposed 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 mark - arrow". Since we know that road marks do not cast shadows, while bright vehicle fragments often do, the "distance to nearest shadow" feature was 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.

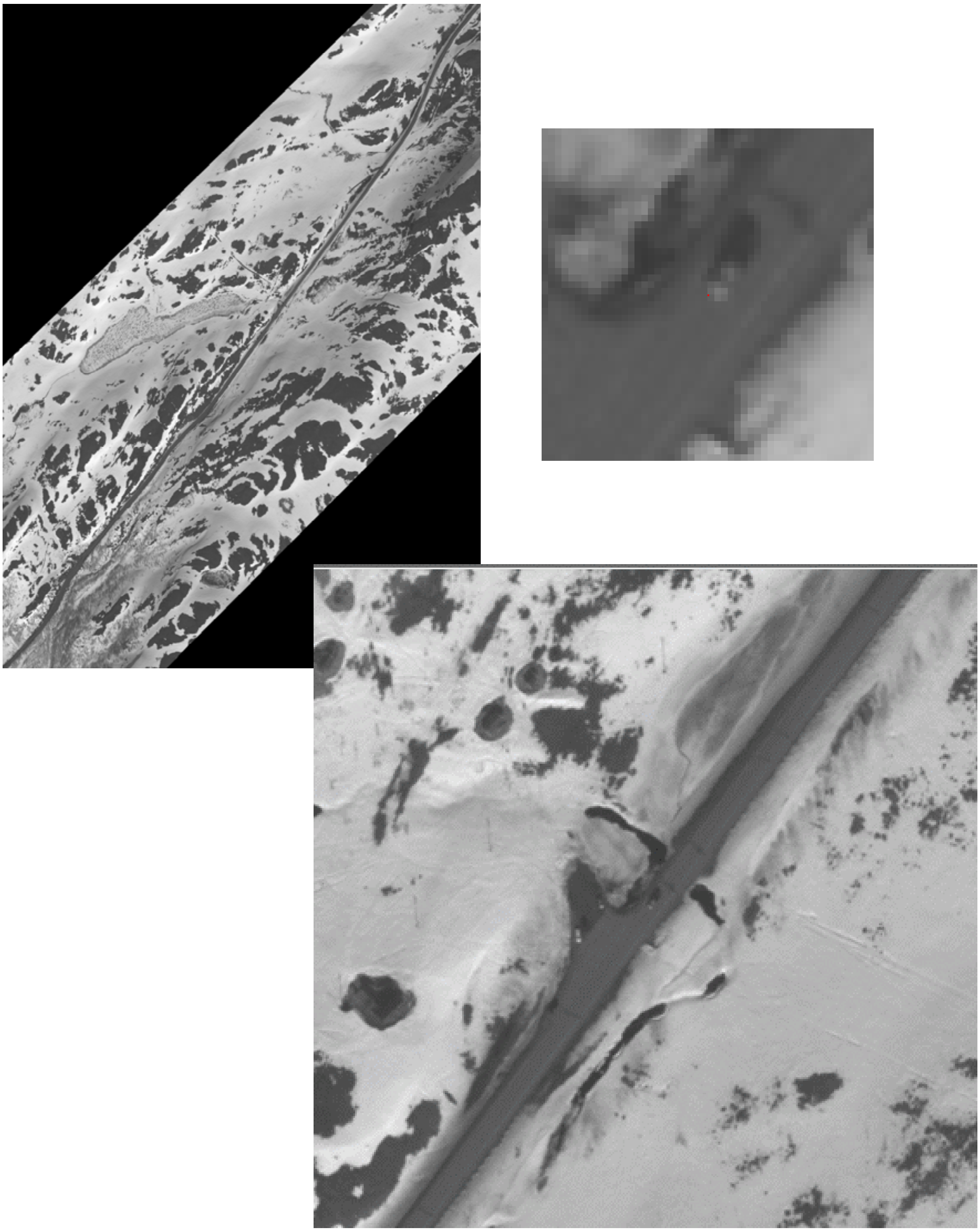


Figure 5.1 A subset of the Sennalandet image at three different zoom levels. Note the vehicle.

Another type of shadows that complicates vehicle detection is shadows by the road, mainly cast shadows from road side vegetation. In section 2.2.2.1 we suggested a simple way of removing such shadows. This approach does not depend on global image properties, such as the shadow mask that was presented in [1].

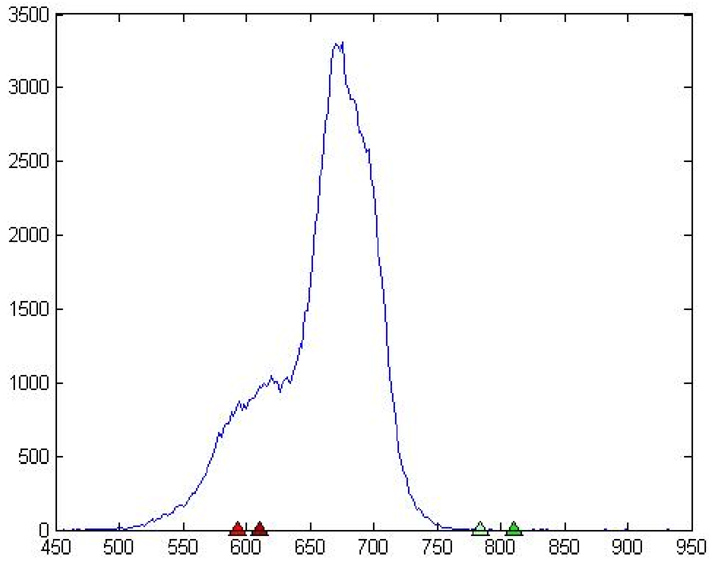



Figure 5.2 Histogram for the Sennalandet image in Figure 5.1. Light green marks the “strict” threshold, which is lower than the dark green “loose” threshold. The histogram has a different shape than the histograms in Figure 2.1.

The segmentation routine was improved by introducing an extra threshold for each stage (bright and dark segments). This reduces the number of vehicles that end up in two parts, while at the same time it avoids the problem of too many segments from using a loose threshold alone.

The selection of features has also been improved from the last version of the algorithm. The choice of features is now based on the results of using an established feature selection method which considers the choice of statistical classifier. The preclassification was significantly modified, based on input from a much larger set of data. The main classification method was improved by introducing class priors.

## 5.3 Recommendations for future work

As mentioned in chapter 4 the satellite image from Sennalandet (in northern Norway) has different lighting conditions from the other cases we have seen, see Figure 5.1. The image histogram shows that the average intensity is higher than usual, and the shape of the histogram is not exactly like described in section 2.2.1, see Figure 5.2 and Figure 2.1. Actually, the “strict” upper threshold is lower than the “loose” upper threshold. It does not make sense to perform hysteresis thresholding with these thresholds. We made a slight modification of the segmentation routine for cases where the “strict” threshold is lower than the “loose” threshold. In such a case we simply skip the hysteresis thresholding, and use the result of the “strict” threshold directly. The algorithm is then able to detect some of the vehicles in the Sennalandet image. However, these observations indicate that further research should be made in order to adapt the segmentation algorithm to different lighting conditions. Furthermore, it is probably wise to have a different set of preclassification parameters for different types of images, and maybe various class descriptions for statistical classification. These concerns also apply to the Eiker image, where the lighting conditions (in part of the image) were the opposite, as a large stretch of road was covered by a cloud shadow, see Figure 5.3.

As described in section 2.3.2 we have used six different classes for classification, and one of the classes is “Bright vehicle fragments”. This was done because we had many samples of vehicles

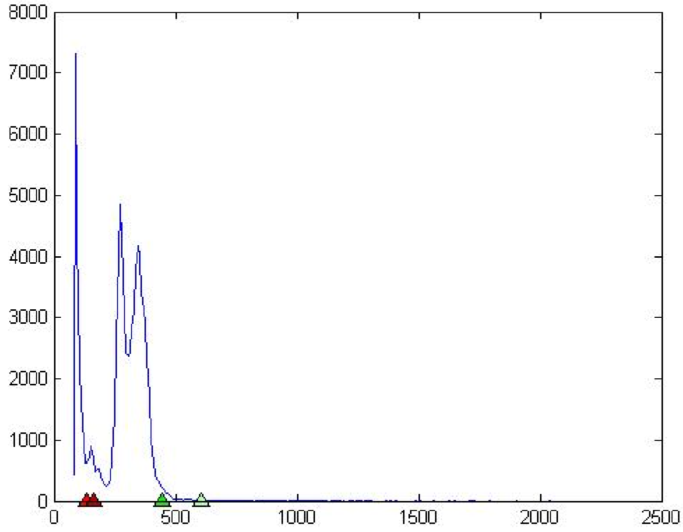



Figure 5.3 Histogram for a part of the Eiker image (on the left side in Figure 4.3).

that had been segmented as two fragments, or only a part (one fragment) of the vehicle was found. After classification we have to connect vehicles that consist of two fragments so that a vehicle is not counted twice. As mentioned earlier, the segmentation step uses two thresholds (a strict and a loose) for each stage of the segmentation in order to avoid disconnected objects.This approach reduces the number of fragmented vehicles. However, the problem is not completely solved, as vehicle fragments are harder to classify than well-defined cars. Especially, vehicle fragments are often confused with road marks. In [4] Hamre suggests a method for connecting fragments. This method is based on the segmentation algorithm that was developed in the "Road Traffic Snapshot" project. Loosely speaking the method looks for other segments within rectangular frames (with different orientations) around each segment. If the size and orientation of the segments meat certain requirements they are connected. The method allows both bright and dark fragments to become connected. This approach shows good results, and may possibly contribute to solve our problem if the right adaptations are made. Alternatively, the information about nearby fragments could be implemented as a (contextual) feature, which again may contribute to improve the classification.

# Acknowledgements

We thank Ragnar Bang Huseby and Jostein Amlien for fruitful discussions in the development of the methodology. Furthermore, Jostein Amlien contributed with manual vehicle countings in the validation experiments.

The work presented in this report was funded by the Norwegian Public Roads Administration (Statens Vegvesen, Vegdirektoratet), and the Norwegian Space Centre (Norsk Romsenter).

# References

[1] L. Aurdal, L. Eikvil, H. Koren, J. U. Hanssen, K. Johansen, and M. Holden, *«Road Traffic Snapshot»*, NR Report no. 1015, Norwegian Computing Center, 2007. ISSN 978-82-539-0525-9.
[2] L. Eikvil, L. Aurdal, and H. Koren, *«Classification-based vehicle detection in high-resolution satellite images»*, submitted to ISPRS Journal of Photogrammetry and Remote Sensing.
[3] The PRTools website, http://www.prtools.org/.
[4] N. Hamre. *«Simulering av QuickBird satellittbilder med egenskapsuttrekking for kjøretøy i oppløsning 0,125-0,6 m, og synbarhet av kjøretøy i SAR flybilder»*, Master thesis, Department of informatics, University of Oslo, Norway, 2008.