Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

SatTrafikk project report 2009

Note no SAMBA/55/09

Authors Siri Øyen Larsen, Arnt-Børre Salberg

Date December 2009

Contract JOP.17.09.2 (Norwegian Space Centre)

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 modelling. 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 project report 2009**
Authors	**Siri Øyen Larsen and Arnt-Børre Salberg**
Date	December
Year	2009
Publication number	SAMBA/55/09

Abstract

Traffic statistics is a key parameter for operation and development of road networks. Vehicle counts based on automated satellite image analysis can provide useful additional information to traditional ground based traffic surveillance. A significant advantage of satellite based technology is that it does not require installation and maintenance of equipment in the road. Moreover, a satellite image can cover large geographical areas, as opposed to traditional ground based traffic measurement equipment.

The SatTrafikk project is a research project conducted by the Norwegian Computing Center and funded by the Norwegian Space Centre and the Norwegian Public Roads Administration, where the main objective is to develop a future system for automated road traffic counts from satellite imagery. The project was started in 2007 as a follow-on to the European Space Agency project "Road Traffic Snapshot" (2006-2007), which demonstrated the feasibiliy of this kind of vehicle detection.

The presented methodology for vehicle detection consists of three main steps: segmentation, feature extraction, and classification. In 2009, we have developed a new approach for segmentation, based on a filtering technique followed by region growing. Some new features to describe the segments have been tested. The classification approach is also new compared to our previous work, and we now propose to use a K-nearest-neighbour classifier with two classes: vehicle and non-vehicle, and to use separate feature sets for bright and dark image objects. The methods were tested on images from Kristiansund, Østerdalen and Sollihøgda showing good results, with an average detection rate of 94.5%, and only 6% false alarms.

The construction of road masks for the satellite images is a pre-requisite for the proposed vehicle detection processing chain. A road mask can be constructed from vector data, provided co-registration of the road vector data with the satellite image. As of today, this registration requires a considerable amount of manual labor. For future operationalization, it is therefore necessary to develop an algorithm for automatic rectification of the road mask to match the satellite image.

Keywords	Remote sensing, pattern recognition, vehicle detection, road traffic statistics, very high resolution satellite images
Target group	Road traffic authorities
Availability	Open
Project number	220 339
Research field	Earth Observation
Number of pages	28

Contents

1 Introduction

Road networks are resources of major importance for the society. Operation and development of road networks is a central activity for several public institutions, such as the Norwegian Public Roads Administration (Statens Vegvesen Vegdirektoratet, SVV). Traffic statistics is a key parameter for this activity. The primary source of traffic statistics today is ground based counts generated using various types of equipment mounted in or close to the road. Such equipment, e.g., induction loops or pressure sensors, counts the number of cars passing a given location on the road during a period of time. Important statistics describing traffic can be derived from these counts, most importantly the so-called AADT, i.e., the Annual Average Daily Traffic, which is the average number of vehicles passing a given location during one day, taken as an average over a year. In Norway AADT is estimated using ground based vehicle counts in combination with statistical tools developed at the Norwegian Computing Center (Norsk Regnesentral, NR).

For fairly large parts of the Norwegian road network AADT is still unknown. The reason is that installation and operation of measurement equipment for ground based counts are both difficult and expensive, hence there are relatively few counting locations as seen in a geographical scope. Especially, AADT is missing for most roads with low traffic density on national basis.

Over the last few years, very high resolution satellite sensors have opened up for alternative means of traffic monitoring. Vehicle counts based on automated satellite image analysis can provide useful additional information to traditional traffic surveillance. A significant advantage of satellite based technology is that it does not require installation of equipment in the road, thus maintenance demands are no longer an issue. Moreover, a satellite image can cover large geographical areas and in principle this allows for AADT estimation of all the roads in the region, as opposed to only a few roads, as one is restricted to using ground based measurements.

In 2006-2007 NR conducted the European Space Agency project "Road Traffic Snapshot" (see http://dup.esrin.esa.it/projects/summaryp92.asp) in cooperation with SVV and Institute of Transport Economics (Transportøkonomisk Institutt). Fundamental algorithms for vehicle detection using satellite images was demonstrated in the project. In the following project "SatTrafikk", begun in the summer of 2007 and funded by SVV and the Norwegian Space Centre (Norsk Romsenter, NRS), the detection algorithm was further developed and optimized. In 2008 the methods were validated on a large data set containing a variety of road conditions from different parts of the country. While the original methods were developed using satellite images covering the urban area of Oslo, this larger data set contained images from Sennalandet, Bodø, Kristiansund, Østerdalen, Eiker and Sollihøgda (see Figure 1.1). These images revealed different types of challenges when it comes to automatic vehicle detection in different parts of the country, and especially there are differences between urban and rural areas.

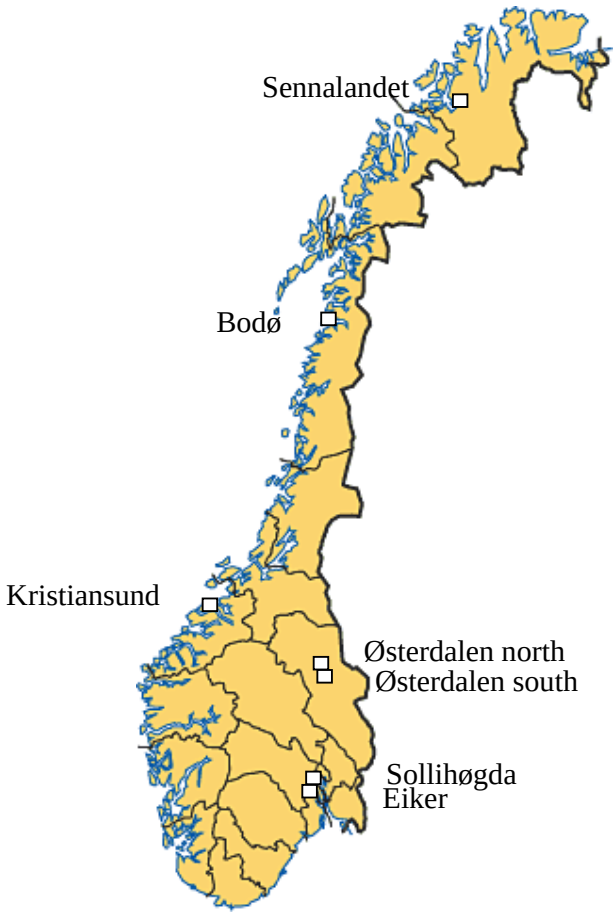

Figure 1.1. Locations of image acquisition for the 2008 QuickBird data set.

In 2008 we also conducted research on how good AADT estimates can be expected using today's statistical model (the basis curve method), given the availability of one or a few satellite images per year. For roads with relatively large AADT as seen in a national context (i.e., AADT >20,000 vehicles) the results were promising (absolute error less than 20% given two satellite images a year), with the precondition that the vehicle detection algorithm is more or less completely accurate. For roads with smaller AADT (<20,000 vehicles) the corresponding average error was around 25%. With AADT less than 1,000 vehicles, even larger error can be expected, although, as traffic statistics hardly exist for such roads, there was no data evidence to verify this.

On the background of the mentioned experiences, the focus of the SatTrafikk project in 2009 has been directed towards relatively low traffic density roads, mainly located in rural areas. Large parts of the previously suggested methods have been replaced or modified to meet the specific challenges related to vehicle detection under these conditions, and to optimize the detection strategy in general. This report desribes the work that has been done to improve the methods, as well as the final algorithm, including experimental results.

2 Satellite data

In order to experimentally test our algorithm we used QuickBird satellite imagery, which has a ground resolution of 0.6m in the panchromatic band. Road masks are applied to the images in order to restrict vehicle detection to roads. The road masks are manually constructed for this study. Areas covered by clouds or heavily influenced by cloud shadows are excluded.

The image data consists of 6 QuickBird satellite images from three different locations in Norway:

- Østerdalen, a long valley in the middle of the country, with only sparse settlemens. Most of the road in this area is located close to forest and has very low traffic density.
- Kristiansund, a city on the north-west coast. The extracted roads are located outside the city in a rural area with a few suburban settlements.
- Sollihøgda, in the south eastern part of the country. A multi-lane highway pass through this area, which has higher traffic density than the other two locations. For experiments, we extracted only a smaller part of this highway, lying in an area where it is reduced to one-lane highway.

Table 2.1 presents an overview of the data. Throughout this report the image ID numbers in the first column of the table will sometimes be used as reference to the image when discussing experimental results. The number of sub images for each scene refer to how many sub images was extracted from the total scene and used for processing. Each sub image contains a road segment of roughly 1-3 km. The extracted road length equals the sum of the road lengths in the processed sub images. To give an impression of the given conditions, some sample image sub sections are shown in Figure 2.1.

Image ID	Location	Road ID	Date (mm.dd. yy)	Time (UTC)	Image area (km2)	Extracted road length (km)	Traffic density (vhcls/km)	Number of sub images
1	Østerdalen	RV3	08.10.04	10:39	59	31.0	1.4	20
2	Østerdalen	RV3	09.06.09	10:29	59	17.1	1.3	13
3	Kristiansund	RV70	06.19.04	10:56	29	5.8	5.7	5
4	Kristiansund	RV70, EV139	07.08.08	10:57	64	11.6	4.1	8
5	Sollihøgda	EV16	05.10.02	10:32	52	3.1	2.9	1
6	Sollihøgda	EV16	08.21.08	10:48	64	3.1	8.4	1

Table 2.1. Satellite image data.

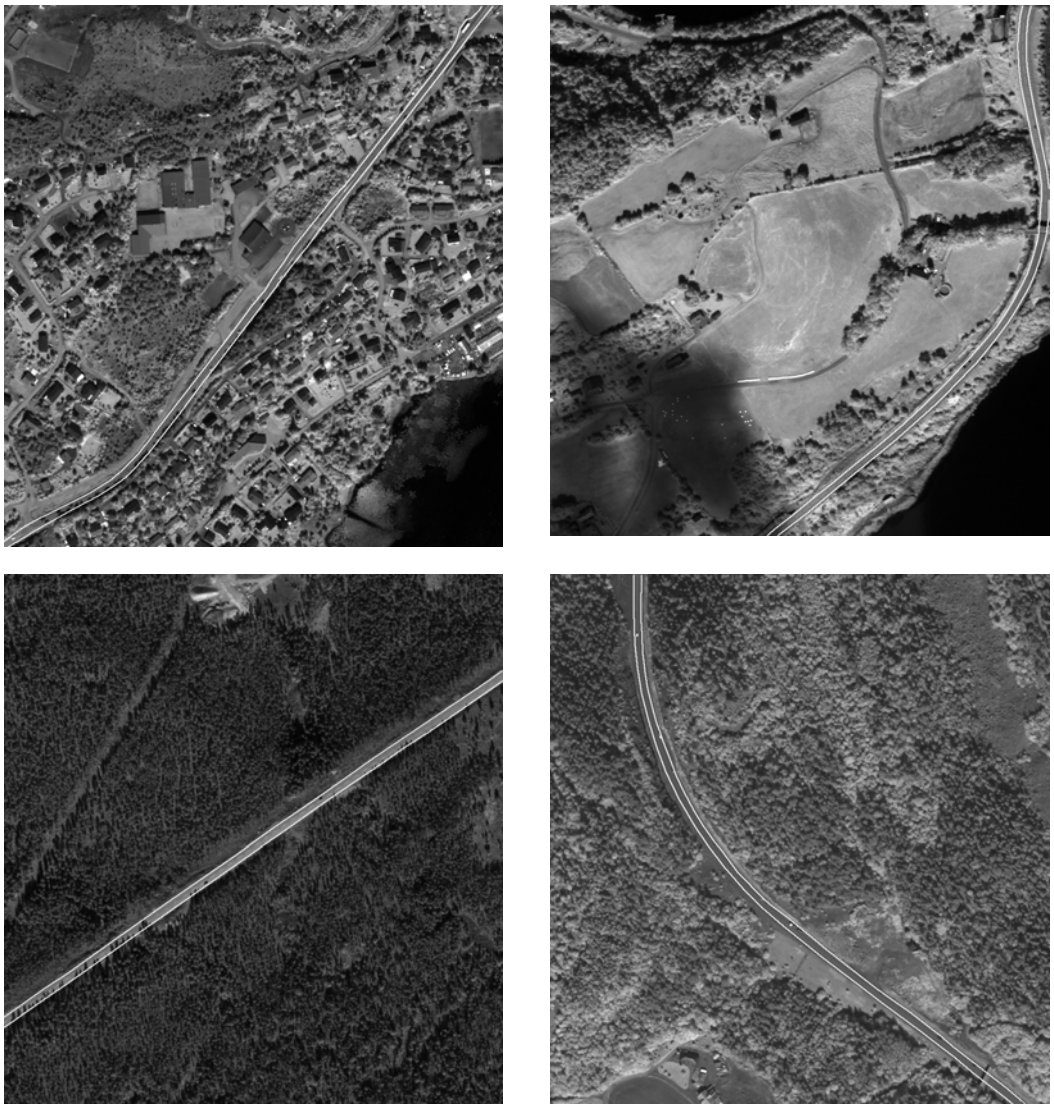

Figure 2.1. These sample images show sections of road extracted from the satellite scenes of (from above left) Krisitansund (image 3, road RV70), Kristiansund (image 4, road EV139), Østerdalen (image 1, road RV3), and Sollihøgda (image 6, road EV16), cf. Table 2.1.

3 Segmentation strategies

Two different segmentation strategies were developed and tested in 2009. The first strategy is related to the thresholding technique used earlier in the project, see [1], while the second is a filtering technique which represents a completely different way of thinking when it comes to segmentation.

3.1 Strategy 1: Morphological contrast enhancement followed by local thresholding

The previously suggested segmentation approach (see [1]) consisted in a modified version of Otsu's method for automatic threshold selection and was based on global thresholding of the image. The success of the approach relied on certain assumptions about the image histogram which did not hold in the general case, and failed to capture local contrast variations.

In 2009, a first attempt to improve the segmentation strategy was the following. Instead of considering the whole image when the threshold is selected, we focus on a small region of the road at the time. This is necessary since we will perform detection in a large area (several kilometers of road), and the local illumination conditions can vary considerably from one location to another within the same image. Furthermore, the local conditions on the road will usually differ throughout the image, due to variations in road surface conditions, road markings, asphalt color, etc. Before thresholding, we perform morphological operations that enhance the visibility of the vehicles. More specifically, we perform gray tone dilation with a structuring element that represents a line, oriented in the same direction as the road, to enhance bright vehicles. Similarly, we perform gray tone erosion to enhance dark vehicles. We then traverse the road along the midline, extract sub images that represent 40-50 m road segments, and assume that the direction of the road is constant within a sub image. We calculate the Otsu threshold for the dilation and erosion results separately. If the separation between the classes exceeds a preset limit, we binarize the image using the Otsu threshold and extract the foreground segments. Tree shadow segments are excluded by rejecting segments that extend outside the road mask. See Figure 3.1 for some examples.

3.2 Strategy 2: Elliptical blob detection using a Laplacian of Gaussian filter

The second segmentation strategy represents a different way of approaching the problem: instead of separating bright and dark intensity levels in the image directly, a small image filter which resembles the kind of objects we are interested in is used to search through the image. Locations where the filter in some sense matches the underlying image are marked as candidate locations for possibly interesting objects. In a second step, object regions must be defined at the locations that were marked during the first filtering step.

3.2.1 Extraction of elliptical blob locations (filtering)

The scale-space representation of an image is a family of successively smoothed images derived from the original image with the intention to represent the original data at multiple scales. It was first introduced by Witkin [3], as an answer to the problem of deciding which neighbourhood size is correct for describing image features. Treating scale as a continuous parameter, a family of smoothed images is derived by smoothing the signal with a mask (filter) of variable size (scale). A common choice is the Gaussian convolution. It appears that features of a given size in the original image become further enhanced at the corresponding scale in scale-space.

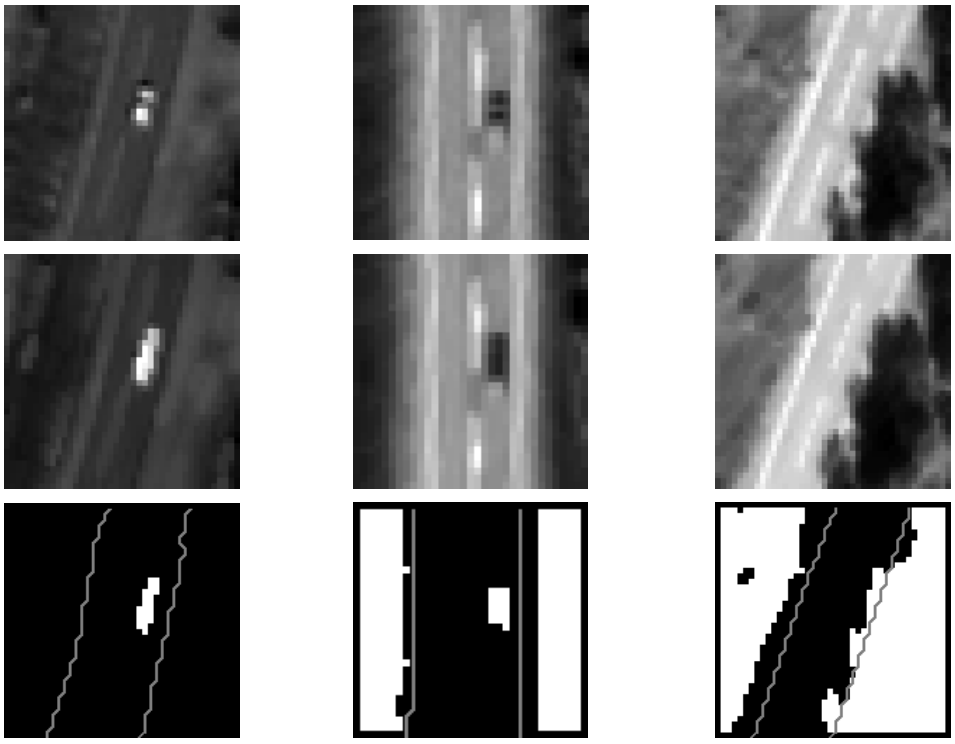

Figure 3.1. The segmentation process illustrated. The columns show three cases of segmentation.Upper row: panchromatic images. Middle row: dilated (left column) and eroded (middle and right column) images. Lower row: segmentation result images. The contour of the road mask is overlaid on the segmentaiton result. The segments (tree shadows) in the right column are discarded since they extend outside the road mask.

The scale-space is formally defined as the solution to the diffusion equation with the original image as initial condition (see e.g. [4]). The solution is equivalent to convolution of the image with the Gaussian kernel

$$g(x, y; t) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right),$$

where σ^2 is the scale parameter, which here equals the squared standard deviation of the Gaussian filter.

Our approach can be seen as a generalization of the disk detection approach presented by Blostein and Ahuja in [5]. They use a region detector derived from properties of a Laplacian of Gaussian ($\nabla^2 G$) scale-space. Circular image regions of uniform gray level are found based on the image response to convolution with first derivative Gaussian filters over a range of scales. The size and location of best fitting disks to such regions are estimated from the observed response of the $\nabla^2 G$ operator across the image and across scales using an analytic expression for the response of $\nabla^2 G$ to disks. They also use a σ-differentiated Laplacian of Gaussian filter ($(\partial/\partial\sigma)\nabla^2 G$) to analyze the behaviour of images across scales. With these two measurements (the image responses to convolution with $\nabla^2 G$ and $(\partial/\partial\sigma)\nabla^2 G$) both disk radius and disk contrast can be estimated, based on derivation of the expected values of these two measurements for an ideal circular disk of constant gray level in the continuous case. Candidate locations for disks in the image correspond to local extrema in scale-space. A candidate blob is accepted only if the diameter is close to the diameter of the center lobe of the filter.

The image features of our interest are vehicles, which are characterized in the image as rectangular regions that are brighter or darker than the local background. However, due to the barely sufficient resolution of the satellite images for this application, the corners of the rectangular vehicles appear smoothed in the image, thus the interesting features can loosely be described as elliptical blobs whose gray level stand out from the surroundings. We find elliptical image regions of uniform gray level by convolving the image with $\nabla^2 G$ filters over a range of scales, and comparing the convolution result to that expected for an ideal elliptical region of constant gray level.

We use an unnormalized, negated Gaussian of two variables, x and y, centered at the origin, and aligned with the coordinate axes

$$G(x, y) = -\exp\left(-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)\right),$$

such that the expression for the $\nabla^2 G$ filter is given by

$$\nabla^2 G(x, y) = \left(\frac{1}{\sigma_x^4}\left(\sigma_x^2 - x^2\right) + \frac{1}{\sigma_y^4}\left(\sigma_y^2 - y^2\right)\right)\exp\left(-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)\right).$$

Equation 3-1

Here, we must define two scale parameters, i.e., the standard deviation in the x and y directions, σ_x and σ_y, respectively.

The $\nabla^2 G$ filter is expected to give a strong positive/negative response at bright/dark elliptical blobs that have roughly the same size and orientation as the filter. We can estimate the contrast C and scale factor S of the elliptical blobs, meaning the numbers C and S such that the ellipse can be written on the form

$$I(x, y) = \begin{cases} C & \text{for } \dfrac{x^2}{\sigma_x^2} + \dfrac{y^2}{\sigma_y^2} \le S^2, \\ 0 & \text{otherwise.} \end{cases}$$

Equation 3-2

Thus, we look for elliptical blobs of a given ellipticity (shape) σ_x/σ_y, but unknown size. In order to estimate both the contrast and the scale of the ellipse, two measurements must be made over the image. In addition to the $\nabla^2 G$ filter, we can use the first derivative with respect to σ, i.e.,

$$\left(\frac{\partial}{\partial\sigma_x} + \frac{\partial}{\partial\sigma_y}\right)\nabla^2 G(x, y) = \left(-\frac{1}{\sigma_x^7}x^4 - \frac{1}{\sigma_y^7}y^4 - \frac{\sigma_x + \sigma_y}{\sigma_x^4\sigma_y^4}x^2y^2 + \frac{5\sigma_y^2 + \sigma_x^2}{\sigma_x^5\sigma_y^2}x^2 + \frac{5\sigma_x^2 + \sigma_y^2}{\sigma_x^2\sigma_y^5}y^2 - \frac{2\left(\sigma_x^3 + \sigma_y^3\right)}{\sigma_x^3\sigma_y^3}\right)\exp\left(-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)\right).$$

Equation 3-3

A closed-form solution to the integrals defining the convolution of an ideal ellipse image (as given in Equation 3-1) with the filters (as given in Equation 3-2 and Equation 3-3) can be derived analytically, using polar coordinates. Performing these calculations it can be shown that at the center of the ellipse $(x,y)=(0,0)$ the responses are given by the following two expressions:

$$\left(\nabla^2 G * I\right) = \pi C S^2 \sigma_x \sigma_y \left(\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2}\right) \exp\left(-\frac{S^2}{2}\right),$$

Equation 3-4

and

$$\left(\left(\frac{\partial}{\partial \sigma_x} + \frac{\partial}{\partial \sigma_y}\right)\nabla^2 G * I\right) = \frac{\pi C S^2}{\sigma_x^2 \sigma_y^2}\left(\left(\frac{3}{4}\left(\sigma_x^3 + \sigma_y^3\right) + \frac{1}{4}\sigma_x \sigma_y \left(\sigma_x + \sigma_y\right)\right)S^2 - 2\left(\sigma_x^3 + \sigma_y^3\right)\right) \exp\left(-\frac{S^2}{2}\right).$$

Equation 3-5

Using Equation 3-4 and Equation 3-5 to solve for the scale parameter S and contrast C we find that

$$S = \sqrt{\frac{4\dfrac{\left(\left(\dfrac{\partial}{\partial \sigma_x} + \dfrac{\partial}{\partial \sigma_y}\right)\nabla^2 G * I\right)}{\nabla^2 G * I}\sigma_x \sigma_y \left(\sigma_x^2 + \sigma_y^2\right) + 8\left(\sigma_x^3 + \sigma_y^3\right)}{3\left(\sigma_x^3 + \sigma_y^3\right) + \sigma_x \sigma_y \left(\sigma_{x+} \sigma_y\right)}},$$

Equation 3-6

and

$$C = \frac{\left(\nabla^2 G * I\right)\sigma_x \sigma_y \exp\left(-\dfrac{S^2}{2}\right)}{\pi S^2 \left(\sigma_x^2 + \sigma_y^2\right)}.$$

Equation 3-7

To summarize, we locate blobs by performing the following steps:

1. Rotate the panchromatic sub image so that the road is (and thus the vehicles are) horizontally oriented in the image.
2. Convolve the rotated image I with the $\nabla^2 G$ and $(\partial/\partial\sigma)\nabla^2 G$ filters defined in Equation 3-1 and Equation 3-2.
3. Find the locations of local extrema in the $\nabla^2 G$ convolution response.
4. Compute the scale factor S and contrast C at the local extrema, using Equation 3-6 and Equation 3-7, and compare the resulting S and C to the corresponding parameters of the filter used for convolution.

Step 3 require some further comments. Since we are going to compare the scale factor of the blob found in the image to the scale factor of the $\nabla^2 G$ filter, we need to know how we can define the latter. The main lobe of the $\nabla^2 G$ filter is the ellipse that satisfies the equation

$$\left(\frac{1}{\sigma_x^4}\left(\sigma_x^2 - x^2\right) + \frac{1}{\sigma_y^4}\left(\sigma_y^2 - y^2\right) \right) = 0.$$

Rewriting this expression to standard form, the main lobe ellipse is given by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \text{ where } a = \frac{\sigma_x}{\sigma_y}\sqrt{\sigma_x^2 + \sigma_y^2} \text{ and } b = \frac{\sigma_y}{\sigma_x}\sqrt{\sigma_x^2 + \sigma_y^2}.$$

Equation 3-8

Here a and b represent the lengths of the principal axes, i.e., the ellipticity a/b is σ_x^2/σ_y^2. However, the formulas for the scale and contrast of an ideal elliptical blob was derived for an image of an ellipse with ellipticity σ_x/σ_y. In order to compare the filter ellipse to the elliptical blob in the image we therefore want to find the σ_x/σ_y-ellipticity ellipse that best fits the main lobe of the filter. This can be expressed as finding the scale $\hat{S}$ which makes an ellipse of the form $\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} = \hat{S}^2$, or, equivalently,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \text{ where } a = \hat{S}\sigma_x \text{ and } b = \hat{S}\sigma_y,$$

Equation 3-9

best fit the ellipse given in Equation 3-8.

This optimization problem can be solved by using the following parametrization for a general ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$:

$$r(\theta; a, b) = \frac{1}{\sqrt{\frac{1}{a^2}\cos^2\theta + \frac{1}{b^2}\sin^2\theta}}, \quad \theta \in [0, 2\pi].$$

The optimal scale $\hat{S}$ is found by minimizing the objective function

$$J(S) = \left\{ \int_0^{2\pi} \left[r\left(\theta; \frac{\sigma_x}{\sigma_y}\sqrt{\sigma_x^2 + \sigma_y^2}, \frac{\sigma_y}{\sigma_x}\sqrt{\sigma_x^2 + \sigma_y^2}\right) - r\left(\theta; S\sigma_x, S\sigma_y\right) \right]^2 d\theta \right\}^{1/2}.$$

This minimization problem can be solved numerically for given values of σ_x and σ_y. The resulting $\hat{S}$ =argmin $J(S)$ will be uniquely determined given the ratio σ_x/σ_y. Now, the scale factor S of a blob candidate can be compared to the corresponding $\check{S}$. In order to make the comparison more intuitive, we actually compare the estimated length of the major principal half axis of the elliptical blob, given by $S\sigma_x$, to the major principal half axis of the σ_x/σ_y-ellipticity ellipse that best fits the main lobe of the filter, given by $\hat{S}\,\sigma_x$, and the blob is kept only in the case that the absolute difference between $S\sigma_x$ and $\hat{S}\,\sigma_x$, is less than 25% of $\hat{S}\,\sigma_x$.

When it comes to the estimated contrast C of the elliptical blob, we require $C>1$. We also apply an amplitude threshold to the $\nabla^2 G$ convolution response in order to avoid local extrema that only represent noise in the data. Furthermore, if there are overlapping blobs, we keep only the blob with the strongest amplitude.

3.2.2 Definition of object regions

Once the blob locations have been marked in the image, we must define the extension of each blob. We use a simple region growing technique, where all neighbouring pixels that pass an intensity threshold are added to the blob region.

This intensity threshold was set based on trial and error. For bright blobs, the best result is obtained using a local threshold, based on the road intensity values in the panchromatic image. More specifically, the bright threshold t_{bright} is defined as

$$t_{bright} = \mu_{loc,\ road\text{-}veg} + 1.5\sigma_{loc,\ road\text{-}veg},$$

where $\mu_{loc,\ road\text{-}veg}$ is the mean intensity of the pixels that are included in the local road mask and at the same time not included in the vegetation and vegetation shadow mask (see Section 4.1), and $\sigma_{loc,\ road\text{-}veg}$ is the standard deviation of the same pixels.

For dark blobs, a suitable threshold in most cases is

$$t_{dark^1} = \mu_{glob,\ road} - \sigma_{glob,\ road},$$

where $\mu_{glob,\ road}$ is the mean intensity of all the pixels in the global road mask and $\sigma_{glob,\ road}$ is the corresponding standard deviation. The exceptions (where t_{dark^1} is not suitable) are local areas where the road cover is darker than usual, e.g., when the road is apparently newly coated with asphalt. We make a test to check if the local mean intensity is darker than usual:

$$\mu_{loc,\ road} < \mu_{glob,\ road} - 0.5\sigma_{glob,\ road},$$

where $\mu_{loc,\ road}$ is the mean intensity of the pixels within the local road mask. In cases where the inequality holds, we apply a local threshold given by

$$t_{dark^2} = (\mu_{loc,\ road} + min_{loc,\ road})/2,$$

where $min_{loc,\ road}$ is the minimum intensity of the pixels within the local road mask.

4 Tree shadows

A frequently encountered problem is tree shadows in the road, cast from vegetation along the side of the road. If we restrict region growing (cf. Section 3.2.2) to the road, we get many tree shadow segments that are easily confused with dark vehicles, depending on the shape of the part of the shadow that lies wihin the road. At the same time, some dark vehicles appear so close to a tree shadow that the vehicle can not be separated from the shadow based on intensity features alone, see Figure 4.1. To solve this dilemma, we do not restrict region growing to the road, but let it also grow in the south-east direction, to see whether it enters a vegetation shadow area. If this happens, the region is checked to find out whether it is a tree shadow, or a combination of vehicle and tree shadow. We will first describe how the vegetation and vegetation mask is defined, and then how we can separate a vehicle from the tree shadows when the vehicle seems to be connected to the shadow.

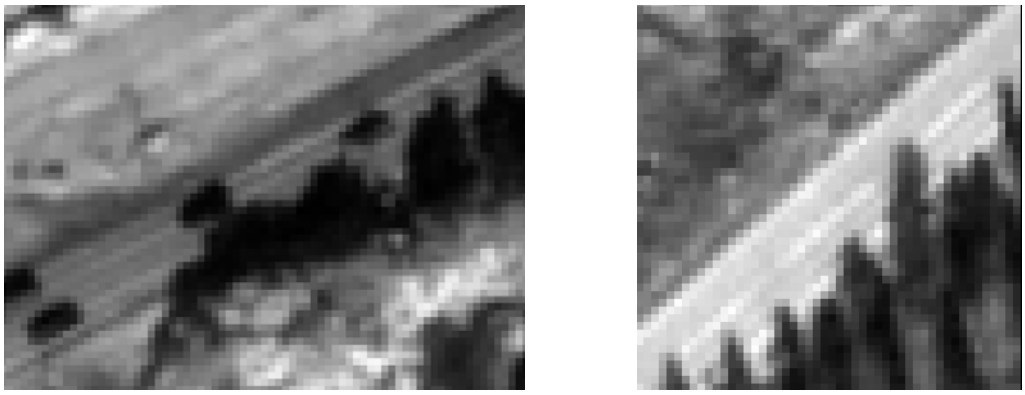

Figure 4.1. Tree shadows in the road. Note that in the left image, one of the dark vehicles is located so close to the tree shadows that it can not be separated from it based on intensity alone. The reason why a human interpreter recognizes the car is the shape of the region.

4.1 Vegetation and vegetation shadow mask

A vegetation mask is constructed from the NDVI (normalized difference vegetation index) image, which is computed from the red and near-infrared bands of the multi spectral image, after resampling to the resolution of the panchromatic image using cubic interpolation. The appropriate theshold to separate vegetation from non-vegetation pixels is found from application of Otsu's algorithm to the NDVI image.

The combined vegetation and vegetation shadow mask is defined by repeatedly dilating the vegetation mask, but for each dilation, only pixels that are below an intensity threshold t_{shadow} are added to the mask. The threshold is given by

$$t_{shadow} = \mu_{glob,\,road} - \sigma_{glob,\,road},$$

where $\mu_{glob,\,road}$ is the mean intensity of all the pixels in the global road mask and $\sigma_{glob,\,road}$ is the corresponding standard deviation (cf. Section 3.2.2). The structuring element used for dilation is:

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

i.e., it represents shadows lying north of the object, since in all the images in our data set, sun light enters the scene from the south/south-east.

4.2 Separation of tree shadows and dark vehicles

When a dark object region appears to be located partly inside and partly outside the road and it overlaps the vegetation shadow mask, we must closely study the shape. The implemented approach can briefly be described as follows. We first extract the contour (border) of the region and estimate its curvature. Then, we start at a point of the contour lying inside the road and traverse the contour in both directions, looking for points of strongly negative curvature, and where the normal direction of the contour agrees with the orientation of the road, i.e., the normal direction of the contour is parallell to the road direction. If points of negative curvature are found close to the road border and in both directions, with opposite contour normal directions, parallell to the road direction, then the region is clipped along the straight line connecting the two points (Figure 4.2). After clipping, the segment inside the road is kept, while the outside part is thrown away. The extracted segments are then carried on to the classficiation stage of the vehicle detection processing chain.

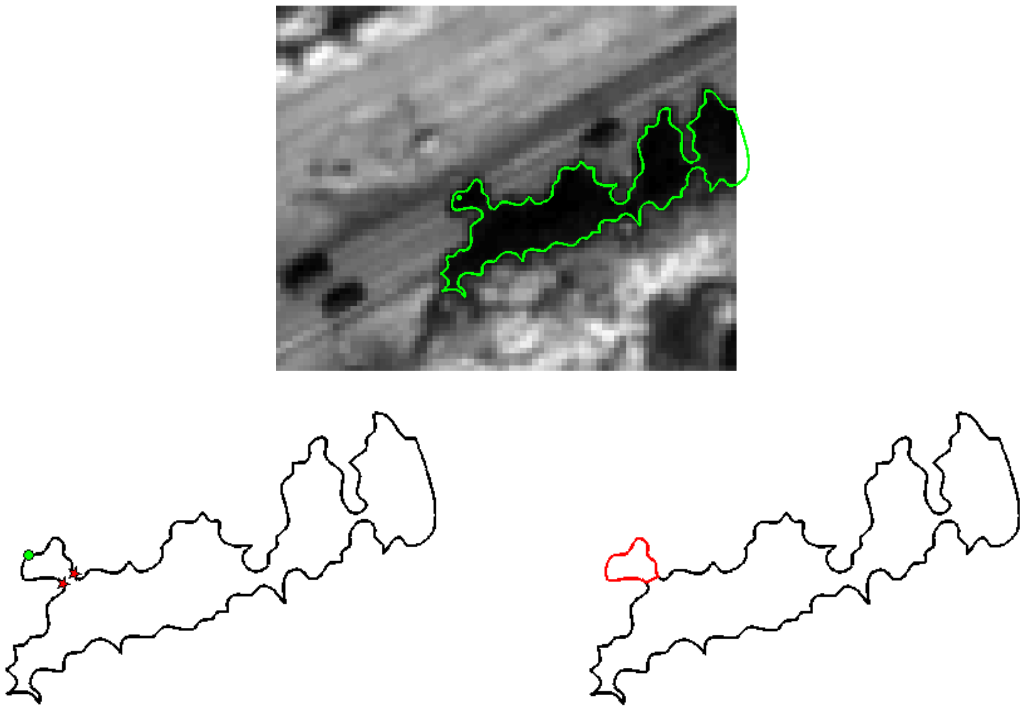

Figure 4.2. Upper image: panchromatic image, overlay: green point=blob center, green outline=contour of the corresponding object region. (The contour has been smoothed using splines). Left below: black contour=contour of the object region, green point=start point for search along the contour, red stars=points with strongly negative curvature and opposite normal directions, both parallell to the road direction. Right below: red contour=contour of the new region.

5 Feature extraction and classification

The whole idea of defining object regions during segmentation is that we can calculate features that describe them – features which assume different values for different kinds of object, and hence, features that can be used to discriminate vehicles from other types of objects.

In 2009, we have followed the same basic approach as previously, in that we extract features and then classify the objects based on some selected feature set. However, some features have been added, and the classification method is new.

5.1 Feature extraction

Using the segmentation strategy described in Section 3.2, some of the features are actually extracted before the region definition, i.e., we save some features related to the blob location. The rest of the features are extracted from the object regions.

5.1.1 Complete list of features

Features extracted from the blob center locations:

1. contrast

   The estimated contrast C, as defined in Equation 3-7.

2. LoG amplitude

   The amplitude of the image response to the $\nabla^2 G$ filter at the scale of the detected blob.

3. σ-LoG amplitude

   The amplitude of the image response to the $(\partial/\partial\sigma)\nabla^2 G$ filter at the scale of the detected blob.

4. estimated major principal half axis

   The estimated size of the major principal half axis of the blob, i.e., the estimated parameter a, when writing the blob ellipse on standard form $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$. More specifically, $a = S\sigma_x$, where S is the estimated scale parameter as defined in Equation 3-6, and σ_x is the parameter used for constructing the filter at the size corresponding to the scale of the detected blob.

5. estimated minor principal half axis

   The estimated parameter b, cf. the previous point of this list, $b = S\sigma_y$.

6. longitudinal contrast 1 and 2

   As explained in Section 3.2.1, the panchromatic image is rotated so that the road is horizontally oriented prior to filtering, i.e., the orientation of the road is parallell to the orientation of the major principal axis of the filter. For a bright blob, the contrast between the blob centers and a point outside the blob ellipse is expected to be positive, and vice versa for a dark blob. We define the contrast in the longitudinal direction as the intensity difference (in the rotated panchromatic image) between the blob center and the point lying 1.5 times the length of the major principal half axis of the main lobe of the filter, to the left/right. See Figure 5.1.

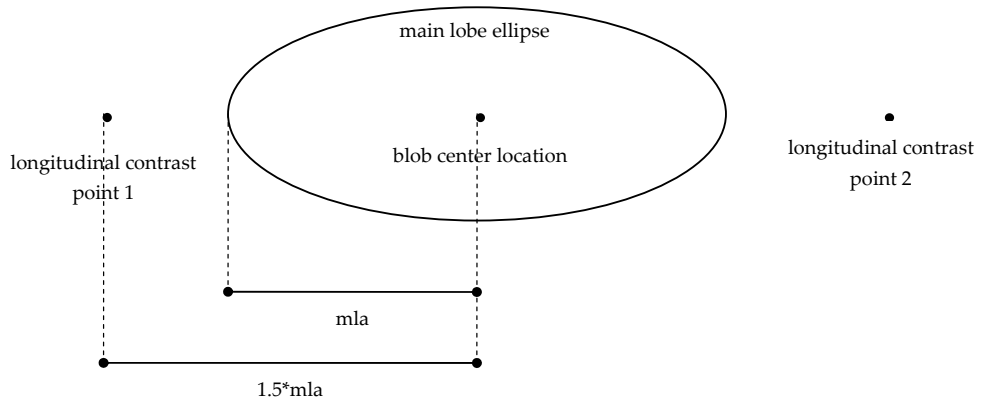

Figure 5.1. Illustration of points used for estimating contrast in the longitudinal direction. The contrast is the difference between the intensity in the blob center location and the longitudinal contrast point 1 or 2. (mla = the length of the major principal half axis of the main lobe ellipse)

Features extracted from the object regions:

7. width

   Width of the bounding box of the object region - after rotation so that the orientation of the region is aligned with the coordinate axes.

8. length

   Length of the bounding box of the object region - after rotation so that the orientation is aligned with the coordinate axes.

9. area

   Number of pixels in the object region.

10. spread

    A measure of the spread of the object region, defined using the region's second order central moments: $(\mu_{20} + \mu_{02})/(\text{area}^2)$

11. area ratio

    The ratio between the area of the object region and the bounding box.

12. elongation

    The ratio between the width and length of the object region.

13. road angle deviation

    The absolute difference between the angle of the road and the orientation of the object region.

14. boundary count

    The number of pixels on the boundary of the region. The boundary is defined using dilation of the object region with a 3x3 flat structuring element.

15. road edge overlap

    The number of object region pixels lying on the boundary of the road mask. The boundary of the road mask is defined using erosion of the road mask with a 3x3 flat structuring element.

16. third order moments

    The third order central moments of the object region: μ_{30}, μ_{03}, μ_{21}, μ_{12}.

17. vegetation distance

    The distance (measured in pixel units) to the vegetation mask measured from the point on the road edge lying closest to the object region centroid.

18. distance from midline

    The distance (measured in pixel units) between the centroid of the object region and the closest point on the midline of the road.

19. distance to road edge

    The distance (measured in pixel units) between the centroid of the object region and the closest point on the road edge.

20. mean pan intensity

    The mean intensity (in the panchromatic image) of the pixels of the object region.

21. pan standard deviation

    The standard deviation of the intensity values of the pixels of the object region.

22. local pan mean

    The mean intensity of the pixels of the object region after local intensity normalization. The local intensity normalization is performed by subtracting the mean and dividing by the standard deviation of the intensities within the road mask, not including pixels in the vegetation and vegetation shadow mask.

23. deviation from global intensity

    The deviation between the mean pan intensity of the object region and the global mean intensity of the road pixels in the image, measured wrt. the global standard deviation of the intensity values.

24. region sobel gradient

    The mean absolute value of the Sobel gradient of the pixels of the object region.

5.1.2 Feature selection

The list above is a complete list of features that were examined. Not all of these features were used for classification in the end. The optimal set of features for classification was selected using so-called feature selection methods, which are designed to loop through various combinations of features and test the performance of some given classifier for each combination. We used some of the feature selection methods integrated in the Pattern Recognition Toolbox PRTools [6] in Matlab.

The feature selection is performed using the entire data set, but for bright and dark objects separately. The criterion (mapping) used for feature selection is the same as the classification method, which is described in Section 5.2. The chosen set of features is:

- for bright objects: contrast, elongation, pan standard deviation, region sobel gradient.
- for dark objects: LoG amplitude, longitudinal contrast 1, length, area, road angle deviation, boundary count, road edge overlap.

5.2 Classification

In the previous stage of the project we used a maximum-likelihood classifier and we assumed that the objects could be grouped into six classes (bright car, dark car, bright truck, bright, vehicle fragment, vehicle shadow, and road mark), each class normally distributed in feature space, with general covariance matrices. This type of classification requires the estimation of parameters for the probability density distributions and the performance relies on whether a high number of training samples is available, not to mention whether the assumption about the type of distribution is true. We experienced that many of the test samples did not belong to any of the six defined classes.

Based on these experienced, we decided to take a different approach to classification. First of all, we use a K-nearest-neighbour classifier, which does not assume anything about the shape of the distribution of the features. Instead, this classifier bases the decision to classify a new sample on how similar the new sample is to samples exisiting in a training set, i.e., a set of samples with known class labels. The K nearest neighbours in the training set each give a "vote" to the corresponding class, and the new sample is classified to the class getting the highest number of votes. The best results were obtained using K=3. Prior to classification, the mean of the feature space is shifted to the origin, and the features are scaled to unit total variance, neglecting class relationships.

Secondly, we split the objects into two sets: bright and dark objects are classified separately. This is done because the features that are best suited to describe bright objects are not the same features that are optimal for dark objects. Thirdly, we use only two classes: vehicle and non-vehicle.

6 Results and discussion

6.1 Segmentation results

The first segmentation strategy was tested on images 1, 3, 4, 5, and 6. The segmentation result was manually inspected and compared to the marked vehicle positions. Based on this inspection we found that 94% of the vehicles were satisfactorily segmented, i.e., 6% were either lost during segmentation, or combined with a non-vehicle object into a joint segment. The total number of segmented objects was approximately eight times that of the total number of vehicles.

The second segmentation strategy was tested on all the images (1 through 6). Based on manual inspection (as in the previous case) we found that 98% of the vehicles were satisfactorily segmented, while 2% were either lost or the region was grown into other objects, resulting in an unrecognizable object. The total number of segmented objects was approximately three times that of the total number of vehicles, thus the number of objects is significantly reduced compared to the first strategy. Leaving out image 2, we can compare the success rate of the second strategy with that of the first. More specifically, going from the first to the second segmentation strategy, the number of successfully segmented vehicles in images 1, 3, 4, 5 and 6 increased from 94% to 97%.

In the remainder of this document we will report results from classification and the total detection chain assuming the second strategy, i.e., the elliptical blob detection approach, was applied for segmentation, since this strategy gave the best results.

6.2 Classification results

Initially, all the objects in all the images were manually labeled as vehicle or non-vehicle, except apparent outlier or doubt objects. Objects that represent car shadows were considered to belong to the vehicle class, as they share similar geometrical and spectral properties as dark vehicle segments. Testing was performed at the sub images (see
Table 2.1), one at the time, leaving the objects from the relevant sub image out of the training set (leave-one-out approach). Weighted by the number of objects in the sub image, this gave a classification error of 0.6% for bright objects and 4.6% for dark objects.

6.3 Final detection results

The final results are found by looking at the number of vehicles reported as detections at the end of the processing chain, and the frequency of false alarms. Since a bright vehicle may be represented by a bright and/or a dark object (the vehicle shadow), the classification output images must be inspected.

The final result image is constructed by adding the bright and dark object regions classified as vehicles, and counting the number of final vehicle objects. To ensure that bright vehicles are not counted twice (the vehicle plus its shadow), bright object regions are dilated in the direction of the expected shadow, i.e., given the known position of the sun in the sky at the moment of image acquisition, in order to ensure overlap of the segments. As pointed out before, the sun angle is approximately the same in all the images in our data set (the sun enters the scene from the south/south-east), and the structure element used for dilation of bright vehicles is:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The final detection results are given in Table 6.1. Intermediate segmentation results are also reported in this table. The detection rate (the fraction of vehicles that are detected) is 94.5%, while the false detection rate (the number of false alarms divided by the number of vehicles) is 6.0%.

Image ID	Location	Vehicles	Correctly segmented vehicles	Correctly detected vehicles	False alarms	Correct detection rate	False detection rate
1	Østerdalen	44	44	43	4	97.7%	9.1%
2	Østerdalen	23	23	23	2	100.0%	8.7%
3	Kristiansund	33	32	30	1	93.8%	3.0%
4	Kristiansund	47	44	42	2	89.4%	4.3%
5	Sollihøgda	9	9	9	0	100.0%	0.0%
6	Sollihøgda	26	26	25	2	96.2%	7.7%
SUM		**182**	**178**	**172**	**11**	**94.5%**	**6.0%**

Table 6.1. Final detection results

6.4 Discussion

As seen in Table 6.1, the detection rate ranges from 89.4% to 100% among the six images. The performance also vary with the location. For example, all the segmentation errors occured in images 3 and 4 - the Kristiansund images. These images contain more clutter than the images from the other locations. The Østerdalen images have more false alarms compared to the number of vehicles than the images from the other two locations. A fair explanation is that the traffic density is lower in Østerdalen. Actually, the average number of false alarms per km is 0.12 in Østerdalen, while it is 0.17 and 0.32 in Kristiansund and Sollihøgda, respectively.

False alarms are caused by e.g. vehicle shadow, trailor wagon (counted in addition to the vehicle pulling it), tree shadow, or spots in the road surface. All the false alarms are shown in Figure 6.1.

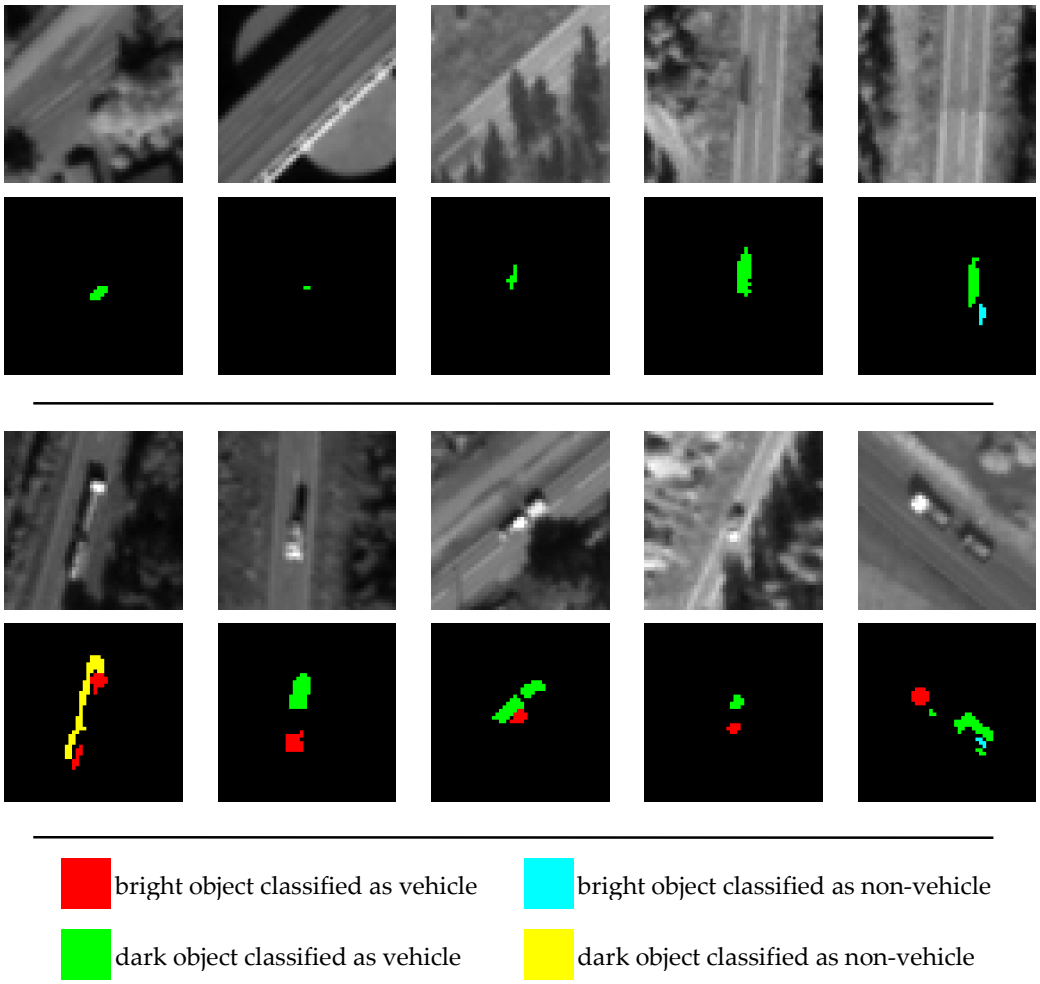

Figure 6.1. False alarms. The classification result is shown below the corresponding panchromatic image, the color of the object region in the classification image indicates which class it was classified to. In each of the examples in the upper row there is one false alarm. Each of the vehicles in the lower row have been counted twice, causing 1 false alarm per vehicle.

There are six false alarms caused by vehicles counted twice, as a result of vehicle shadows or trailor wagons being counted in addition to the actual vehicle.

Two false alarms are caused by tree shadows – one of which represent a tree shadow region whose shape resembles that of a vehicle. The fact that there are so few errors caused by tree shadows is a significant and important improvement compared to previous results.

The cases where segmentation failed (using the second strategy) are shown in Figure 6.2. As can be seen here, three out of four failed segmentation examples are caused by the object region growing too big. In only one case, there is no object at all, meaning, no blob was located at this vehicle location.

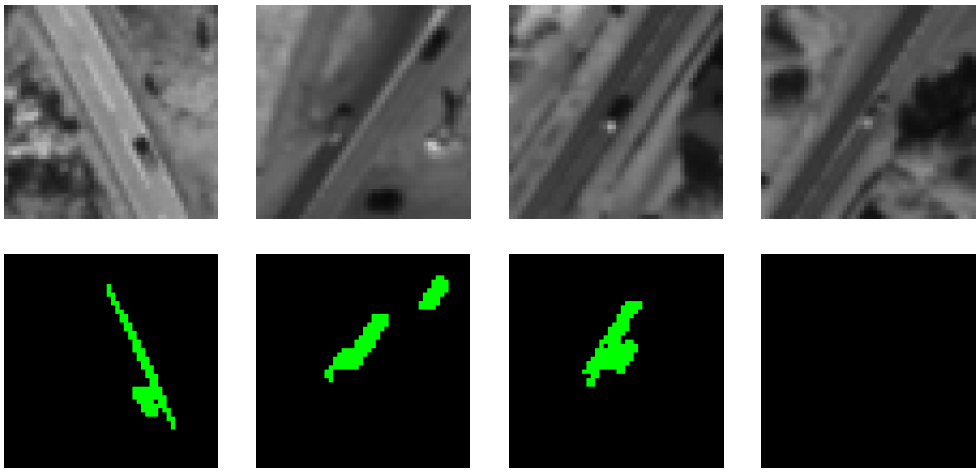

Figure 6.2. Failed segmentation examples. The segmentation result is shown below the corresponding panchromatic image. Dark object regions are colored green. In the first three examples, the region growing failed, and the resulting object is unrecognizable. In the last example, no blob was found at all. The first example is from image 3, the three remaining examples are from image 4 (thus, all examples are from the Kristiandsund scenes).

All the vehicles that were missed becaused of classification error are shown in Figure 6.3. The misclassified dark vehicles in these images typically represent "ugly" object regions, i.e., it can be discussed whether it is actually the region growing step of the segmentation routine that failed, and not the classification. In one case, a bright vehicle with poor contrast to the local background was misclassified.

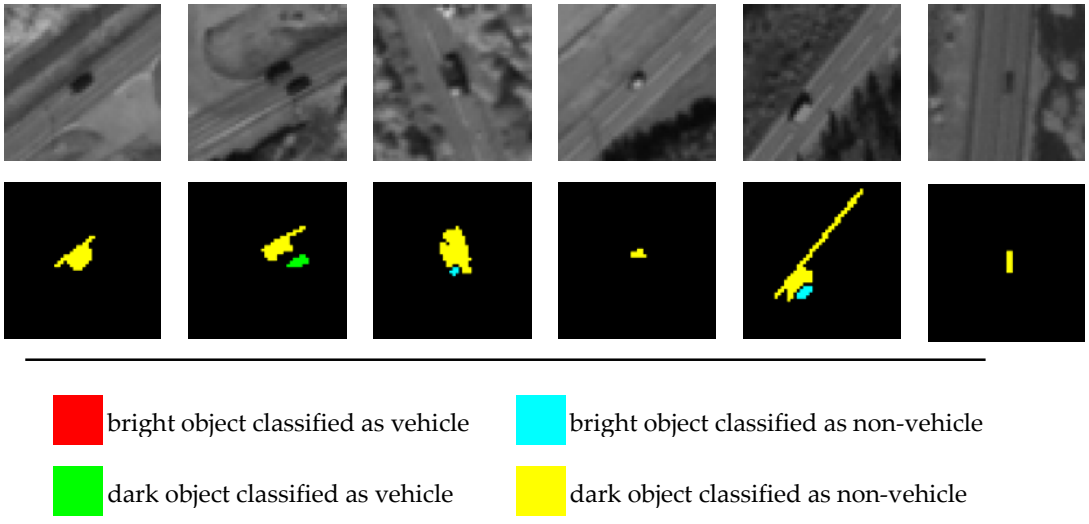

Figure 6.3. Classification errors. The classification result is shown below the corresponding panchromatic image, the color of the object region in the classification image indicates which class it was classified to.

7 Summary and conclusions

In this report we have presented our latest approach for vehicle detection using very high resolution satellite imagery. The development of this methodology is the main activity of the project SatTrafikk, which has been running for three consequtive years now, and is funded by the Norwegian Space Centre and the Norwegian Public Roads Administration. In 2009, we have achieved major improvements concerning vehicle detection rates. We have focused the attention to smaller highways representing typical Norwegian road conditions, i.e., relatively narrow roads, low traffic density, located in mostly rural areas, where roads are often partially covered by tree shadows, as seen from the satellite. The processing chain starts with a panchromatic satellite image and a corresponding road mask, and consists of the steps segmentation, feature extraction and classification, as before. However, especially the segmentation and classification stratgies have been changed compared to our previous work.

The proposed segmentation strategy is based on a so-called Laplacian of Gaussian filter which is used to search through the image for elliptically shaped "blobs", i.e., regions of relatively constant intensity that is brighter or darker than the local background. The size of the elliptical main lobe of the filter corresponds to the size of typical vehicles, and the orientation corresponds to the orientation of the road, hence the vehicles, in the image. Interesting locations in the image can be extracted from the image response to the filter. Although this approach is robust towards local contrast changes, and extract nearly all the vehicle positions in the image, it also finds many candidates representing other kinds of objects. We are able to get rid of most of the non-vehicle candidates using estimates of blob size and contrast, derived from mathematical formulas. The remaining candidates are sent to the feature extraction and classification steps, after a region growing step where the extention of the object region (initialized at the blob candidate location) is defined using a region growing technique.

Variuos features are used to describe the object regions, which are separated into two sets: bright and dark objects. The features describe the intensity, shape and context of the objects. Next, the objects are classified using a K-nearest-neighbour approach (with K=3). This classification approach is different from the previously applied maximum-likelihood classifier in that it does not assume anything about the shape of the distribution of features within a class, and does not require the estimation of parameters.

The entire processing chain was validated using QuickBird images from three different locations (Østerdalen, Kristiansund and Sollihøgda) in Norway (two images from each location). Comparing the result to the manually marked vehicle positions in the images, we found that 94.5% of the vehicles were detected, and the false alarm rate was 6%.

Compared to our previous studies, the detection rates have been significantly improved, and may in many cases now be considered acceptable for operational use. However, there are still some aspects that should be adressed.

First of all, the approach to handle tree shadows is new, and may need some adjustments. It should therefore be validated on a larger set of images.

Secondly, false alarms due to double count of the same vehicle should be avoided. The vehicles should be classified into groups based on size, e.g., car, van, and truck/bus/trailor wagon. Objects regions that are located close to eachother must be seen in context to determine whether they belong to the same vehicle.

Third, image areas containing heavy clouds or cloud shadows have not been used for algorithm development so far. Although clouds prevent optical imaging, it may still be possible to get information from areas affected by cloud shadows. However, the method must be adjusted since the contrast is very different in these areas. Furthermore, the localization of clouds and cloud shadows in the image should be performed automatically, hence, methods for construction of cloud masks must be developed.

Finally, for operational use, the roads must be automatically localized in the satellite image. The position of the mid line of the road is available as vector data together with rough estimates of road width. However, in order to construct a road mask, these data must be co-registered with the satellite image. As of today, this requires a considerable amount of manual labor. It is therefore necessary to develop algorithms for automatic rectification of the road mask to match the satellite image.

References

[1] Larsen, S. Ø., H. Koren, and R. Solberg (2008). SatTrafikk – Vehicle Detection in Satellite Images for Development of Traffic Statistics, NR Note no SAMBA/18/08, Norwegian Computing Center.

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

[3] Witkin, A. P. (1983). Scale-space filtering. In *Proceedings of 8th International Joint Conference on Artificial Intelligence*, Karlsruhe, Germany, August 1983, pp. 1019-1022.

[4] Lindeberg, T. (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2), pp. 79-116.

[5] Blostein, D. and N. Ahuja (1989). A multiscale region detector. *Computer Vision, Graphics, and Image Processing*, 45, pp. 22-41.

[6] Duin, R.P.W., P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, D.M.J. Tax, and S. Verzakov (2007). PRTools4.1, A Matlab Toolbox for Pattern Recognition, Delft University of Technology. http://www.prtools.org.