Operational traffic monitoring using very high resolution satellite imagery

Siri Øyen Larsen[1], Arnt-Børre Salberg[1], Line Eikvil[1], and Øivind Due Trier[1]
Norwegian Computing Center, Oslo, Norway; siri@nr.no

Abstract. In this paper we present a solution for the extraction of vehicle counts from very high resolution satellite imagery intended for use in an operational setting, in order to accommodate road traffic authorities with traffic statistics. In the proposed system, we have developed separate, fully automatic methods for each of the necessary steps: road detection, cloud and cloud shadow detection, and finally, vehicle detection and classification. Input to the system is the satellite image and road vectors, containing geographical coordinates, including height information. The proposed system for extracting vehicle counts has been successfully demonstrated on a collection of WorldView-2 and Quickbird images from Norwegian rural area roads. The true detection rate is on average 85.4%, and false detection rate is 8.6%, i.e., the reported number of vehicles is on average 94% of the correct number. A key parameter to the road authorities is the annual average daily traffic (AADT), which can be estimated based on short-term counts of traffic. The average error in AADT was estimated to 25% for low traffic roads (AADT<20,000 vehicles) using only two satellite images a year. The system will now be evaluated and implemented at the Norwegian Road Authorities as a complementary system for extracting traffic statistics.

Keywords. Vehicle counting, traffic statistics, road detection, cloud detection.

1. Introduction

The increasing availability of very high resolution remote sensing imagery has opened new opportunities for road traffic monitoring applications. While many traffic authorities struggle to create traffic statistics for as many roads as possible with traditional mobile ground based equipment for traffic counts, vehicle detection from satellite images has a potential to cover large geographical areas and can provide valuable additional information to the existing methods. This is especially the case for roads in rural areas and roads with sparse traffic, where ground based measurements are seldom prioritized due to high costs, limited personnel, or related difficulties.

Our objective is to develop an operational solution for the extraction of vehicle counts from very high resolution satellite imagery. In the proposed system, we have developed separate, fully automatic methods for each of the necessary steps: road detection, cloud and cloud shadow detection, and finally, vehicle detection and classification. Input to the system is a satellite image and road vectors, containing geographical coordinates, including height information.

Current commercially available satellite sensors have panchromatic resolution down to 0.5 m. Even at this very high resolution, substructures of vehicles (having an average size of about 2m×5 m) can rarely be distinguished, and explicit models, as they are mostly used for vehicle extraction from aerial images, cannot be used [1]. Zheng *et al*. [2] and Jin and Davis [3] have presented morphological shared-weight neural networks for vehicle detection in satellite imagery of city scenes. Gerhardinger *et al*. [4] apply procedures based on the automatic classification of objects, using an iterative learning approach based on spectral signature and spatial context analysis, to derive city traffic counts from images of Baghdad, using a manually digitized road surface layer on sample zones of the city. They point out that good vehicle classification results are dependent on precise vector data describing the road surfaces, and that automatic procedures for extraction of

road surface should be developed if the work should have to be conducted on larger areas (like the entire city). Sharma *et al.* [5] present a pixel-based Bayesian background transformation approach, showing good performance in Ikonos images of central Ohio, but it strongly depends on the existence of a high quality estimate of a background image.

Most of the studies found in the literature focus on highways or heavily trafficked city roads, and are demonstrated on roads within a rather limited geographical area [1]-[5]. Our approach has a different focus, in that vehicle detection takes place in entire satellite image scenes, containing rural (non-urban) areas, relatively narrow roads (two lanes only), sparse traffic, varying atmospheric conditions (including partial cloud cover, haze and shadow), as well as low sun angle. The analyzed images typically cover an area of 50-100 km^2, and the road to be analyzed 30-40 km. Moreover, the road view is often partially blocked by vegetation (tree crowns) and shadows cast from trees located on either side of the road. We have also focused on developing methods that require no manual interaction at all, as the final system is intended for use in an operational setting, deriving traffic statistics for the national (Norwegian) road traffic authorities.

2. Methods

The proposed operational system consists of three detection modules: road detection, cloud and cloud shadow detection, and vehicle detection. The detection results are the number of vehichles and the cloud-free road length, which are used in a calculation module to estimate the number of vehicles on pre-defined road segments per hour. Each of the detection methods are described in the next sections.

2.1. Automatic road segmentation

The objective of the road detection module is to define a road mask that corresponds to the location of the road in the panchromatic image, so that vehicle detection can be restricted to those areas. Vector data are available from a GIS and need to be exploited both to identify the correct roads, as there may be several roads covered by an image, and to find the road positions in the images. If the images and vector data were very accurately co-registered, the road could be delimited simply by selecting an area corresponding to the width of the road around the vector data. Unfortunately, this is not the case (Figure 1, left). There are two reasons for this: (1) the current national digital elevation model of Norway is not detailed enough to allow for accurate automatic georeferencing of very high resolution satellite imagery, and (2) manual georeferencing should be avoided in a system which is intended to be otherwise fully automated. However, the vector data still provide useful information about the approximate position and trace of the road in the image.

Our method starts by extracting an image area around the road vector, with a size relative to the expected magnitude of the geographical displacement. This area is found by sampling the image along lines perpendicular to the road vectors. The result is a long and narrow image along the road. In this way, the complexity of the road trace is heavily reduced. (In case of perfectly co-registered vector and image, the road would run as a straight line along the middle of the transformed image). In the transformed image, we know that the road must run through all the lines, hence, we analyze the image line by line, searching to trace the road. We perform this search using a snake-based approach, initialized by the road vector. A similar method has been used in the past [6], but without our transformed image space, which significantly simplifies the process. The snake is guided by external and internal forces, which are determined by the image data and vector data, respectively. More specifically, the external force is found from the transformed image, using the blue band of the multi-spectral image. The blue band was found to give the best contrast between road and surrounding landscape, based on visual inspection. The internal force is based on local smoothness

of the trace, where the assumption is that the position will not change dramatically over adjacent lines. We have also developed an adaptive weighting system of the two terms, such that the external force is zero in areas covered by clouds. We then use dynamic programming (more specifically, the Viterbi algorithm [7]) to find the set of points that optimize the sum of internal and external forces over the total set of lines. Since we are using the blue band up to this point, the resulting road mask must finally be resampled and adjusted in order to achieve a smooth representation in panchromatic resolution. We apply a region growing step, initiated by the resampled road mid line, followed by a series of morphological operations and distance thresholding, to obtain the final result (Figure 1, right).

Figure 1: Left: road vector (red contour) directly overlaid on the (here: pansharpened) image. Right: resulting road mask (red layer) overlaid on panchromatic image.

2.2. *Cloud and cloud shadow detection*

When using cloud contaminated images, cloud and cloud shadow masks are required to assist the detection of road and vehicles, and to estimate the correct observed road length for the statistics. We apply a classification based approach, classifying the image into clouds, cloud shadows, green vegetation, water, haze, and bare ground. We model the data representing each class using a multivariate Gaussian distribution where the mean vector and covariance matrix is estimated from the training data, which is constructed by visual inspection and labeling of regions in a set of training images.

An important challenge in this type of classification problem is that there may be a poor match between training data and test data, due to atmospheric, geographic, botanic and phenological variations of the image data. To solve this we build on earlier approaches [8], [9] that aim at exploiting the intrinsic relationships between the training and test data. Although the data distribution for a given class varies between the training images, and also varies between the

3

training images and the test images, the training data domain and the test data domain are generally neither identical nor uncorrelated. This makes it possible to adapt the training data distributions to the corresponding distributions in the test domain.

Bruzzone and Prieto [8] propose a method for retraining the classifier when the test data differs slightly from the acquired training data. When applying this method to our data we obtain a very good statistical fit of the likelihood to the test image, but the mixture components have no longer a physical meaning in the sense that, e.g., the mixture component of a given land cover type no longer models that land cover type, but something else. Building on this method we have therefore developed an alternative approach which applies a low rank modeling of the parameters in order to reduce the number of degrees of freedom and the flexibility of the model [10]. In this way we force the class structure of the training data to be maintained in the test image. We also extend it by incorporating several training images, each with different class dependent data distributions.

For classification of clouds and cloud shadows we use the multispectral image, which first has been downsampled by a factor of eight. We organize the classification process in two stages, where we in the first stage classify the clouds, and in the second stage classify the cloud shadows [10]. The detected cloud pixels are masked in the test images prior to the cloud shadow classification stage. For classification of clouds, we use bands 2 and 3 (red and green bands), as features in a Gaussian distribution. We assume that the class covariance matrix in the test image is the average of all the training image covariance matrices. The mean vector is assumed to be the average mean vector, plus a component of rank one. For cloud shadow detection we have also included an NDVI band and the ratio between bands 2 and 4 (the green and the infrared bands) as features. Here we model a given class mean vector in the test image as a weighted average of the corresponding training image mean vectors, constrained to have only non-negative weights.

Figure 2: Result of cloud and cloud shadow classification. Clouds and cloud shadows are outlined in red and blue, respectively.

To avoid false positive cloud shadow areas, we have added a postprocessing step, where we apply context information about the azimuth and elevation angles of the satellite and the sun. The shadow of each cloud is located in the opposite direction relative to the sun apparent azimuth. However, the projected position of the cloud in the image is different than its actual geographical position for off-nadir imaging. For each cloud, the corresponding shadow is searched for by finding

the shadow direction relative to the projected cloud position. This direction is independent of the cloud height and may be calculated explicitly [11].

2.3. Vehicle detection

The vehicle detection uses a two-stage object segmentation approach, followed by classification, and, finally, object linking (for vehicles consisting of more than one object). In the first segmentation step, 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.

As seen from above, vehicles have rectangular shape, but at ~0.5 m resolution, the corners are blurred, and elliptical filters seem to be a good choice to search for vehicles in the image. We have extended the scale space circular blob detection approach proposed by Blostein and Ahuja [12] to the more general approach of detecting elliptical blobs. We find bright or dark elliptical blobs in the image by convolving it with Laplacian of Gaussian filters over a range of scales, and comparing the result to that expected for an ideal elliptical region of constant gray level. Details of this step are described in [13].

Once the blob locations have been marked in the image, we use region growing to define the shape of the object. Starting from the center pixel of the detected blob, we include neighboring pixels according to certain intensity thresholds that are defined automatically based on local values. From each object we then extract a number of features that can be used to separate vehicles from other type of objects, such as road marks, tree shadows, road surface patches, etc. The features used for classification include spectral, geometrical and contextual features, and the optimal feature set is found for bright and dark blobs separately. Classification is performed using k-nearest-neighbors (kNN), after shifting the feature space to the origin, and scaling the features to unit total variance.

To allow for efficient construction and updating of a training data base, we keep only one non-vehicle class, while the remaining objects are identified as belonging to one of the following six classes: 1) passenger car, 2) truck (or heavy vehicles), 3) bright spot on dark car (reflection of the sun, often seen as overexposed spot in the image), 4) motorcycle, 5) small vehicle shadow (shadow
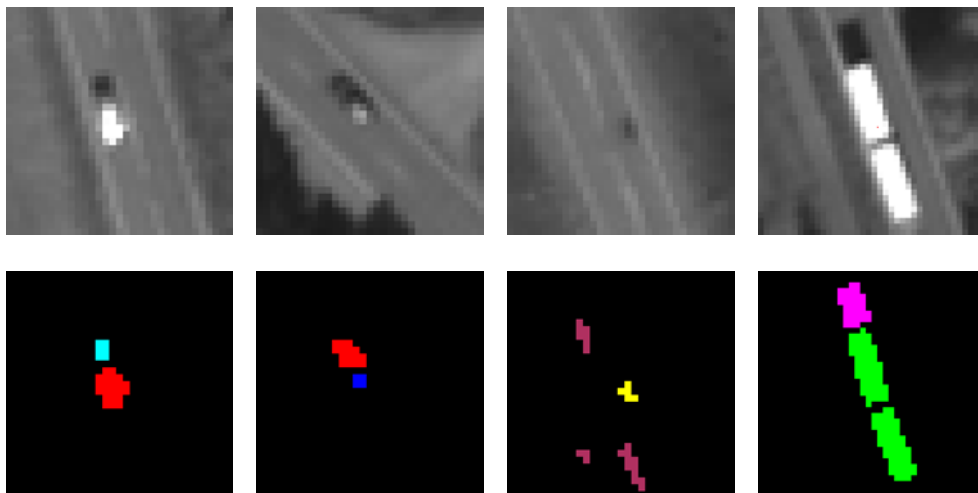

Figure 3: Class division scheme for classification of road objects. The upper row shows example objects in original image. The second row shows the corresponding segmentation result. Color codes: red - passenger car, green - truck, yellow - motorcycle, blue - bright spot on dark car, cyan - small vehicle shadow, magenta - large vehicle shadow, maroon - other (non-vehicle).

5

of passenger car), 6) large vehicle shadow (shadow of truck), see Figure 3. After classification, each vehicle may be represented by one or more road objects - typically, the vehicle and its shadow, but we also have several examples where either the vehicle or the shadow object is lost, either in the segmentation or the classification. Therefore the objects should be linked, in order to ensure that each vehicle is counted once, and only once. For each object classified as vehicle or vehicle shadow, we construct a rectangular box with the same center, width and length as the object, but oriented in the same direction as the road at this location. We then extend the box in the length direction, with the idea that overlapping boxes may be treated as one vehicle. Finally, the vehicle types (size classes) are registered. If a vehicle consists of objects with non-consistent vehicle types (e.g., passenger car+large vehicle shadow, or passenger car+truck), we choose the largest type.

2.4. *Estimation of number of vehicles per road segment*

The national road database in Norway contains a number of road segments, with one segment typically extending from one intersection to the next. For the purpose of vehicle counting, one may assume that no vehicles enter or leave the road in the middle of a road segment. The number of vehicles per hour for a given road segment can then be computed from the number of detected vehicles within the segment, the observed length of the segment within the satellite image, and the speed limit on the road segment, as follows:

number of detected vehicles · vehicle speed (km/h) / length of observed road (km.)

In the current situation, the Norwegian Public Roads Administration uses ground-based sensors to collect traffic statistics. The basis curve method [14] is used to estimate the annual average daily traffic (AADT), including precision estimate, for count sites where counts are available for only a limited part of the year. The method is based on a statistical regression model, the complexity of which is adapted to the amount of data available. Traffic statistics have historically been collected for a large number of roads in Norway, and given us an overview of statistical trends (Figure 4). Given these trends, and the date and time of image acquisition, it is possible to estimate the yearly average number of vehicles per day for the imaged road segment.
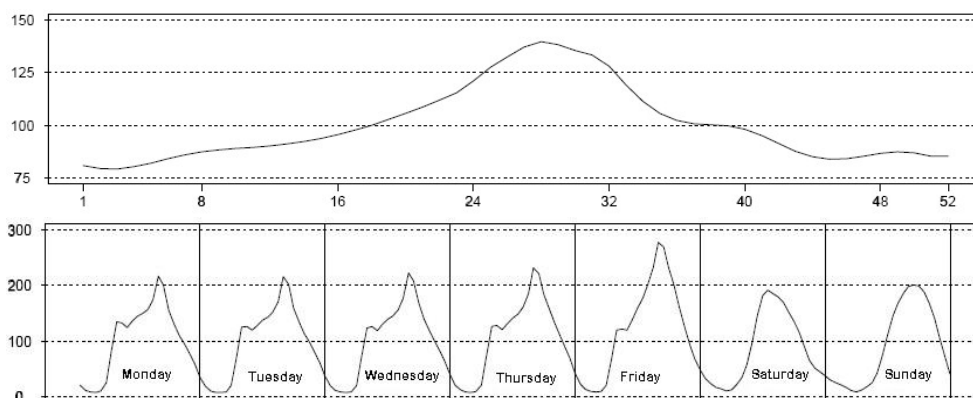

Figure 4: The above graphs show normal variation in traffic volume for small-town/countryside roads and totally for all vehicle size classes on weekly over a year (top) and hourly over a week (bottom) scale.

Assuming that satellite data may be used for reliable traffic counts, there would still be a very restricted amount of data available for input to the basis curve method. Detecting vehicles from space requires very high resolution imagery, taken on a cloud free day, preferably during the snow free season of the year. The elevation of the sun in the sky also plays an important role for the quality of the images. With the commercial satellites of today, this means that we could expect from one to a few useful images per year. One image covering a five kilometer road section where the average speed is 60 kilometres per hour could hold the information equivalent to five minutes long in-road (i.e., a single point in the road) count. It is important for the road authorities to understand how reliable the estimate for the amount of traffic is, based on one single image acquisition, and based on multiple acquisitions. This was the purpose of a separate study [15] where we analyzed whether a few counts of only a few minutes each is sufficient for the basis curve method to make an acceptable estimate of AADT. 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 fairly accurate. For roads with smaller AADT (<20,000 vehicles) the corresponding average error was around 25%. With AADT less than 1,000 vehicles, a larger error is expected, although, as traffic statistics hardly exist for such roads, there was no data evidence to verify this.

3. Results

The methods were validated on a total of five QuickBird and WorldView-2 scenes (Table 1). In each case, the entire road in the image was analyzed, except those parts where the road has more than one lane in each direction (according to manually constructed "ignore area" mask), and areas covered by clouds (as defined by the automatically constructed cloud mask).

Table 1. Experimental data set.

Location	Lon.	Lat.	Route	Date	Time (UTC)	Sun elev.	Image area (km^2)	Obs. road len. (km)
Østerdalen north	10.8	62.0	Rv3	Aug 10, 2004	10:39	43.1	59	31.6
Østerdalen south	10.8	61.7	Rv3	Aug 10, 2004	10:39	43.4	94	43.1
Østerdalen north	10.8	62.0	Rv3	Sep 6, 2009	10:29	33.9	59	28.8
Mosjøen	13.1	66.1	E6	Jun 10, 2010	11:30	47.0	95	31.0
Nordkjosbotn	18.9	69.2	E6	Jun 8, 2010	11:02	43.8	55	12.7

The validation was performed manually as follows: for each scene we made a "ground truth database" of vehicles that should be found. Then, the algorithm was applied, and for each vehicle in the ground truth database, the operator registered 1) whether or not it was detected, 2) whether or not the vehicle type registration was correct, and 3) whether or not it was counted twice. The remaining detections were registered as false. Results are presented in Tables 2 and 3. The true detection rate is 85.4%, and false detection rate is 8.6%, i.e., the reported number of vehicles is on average 94% of the correct number. False detections include road ditches, road marks and asphalt patches (Figure 5). Missed detections include vehicles with low contrast to the road, vehicles overlapping tree shadows, and low contrast due to cloud shadows or haze (Figure 6).

7

Table 2. Detection rates per image.

	Østerdalen north 2004	Østerdalen south 2004	Østerdalen north 2009	Mosjøen 2010	Nordkjos-botn 2010
true vehicles (manual count)	43	129	39	71	32
system detection rate (%)	109.3	99.2	92.3	81.7	75.0
true detection rate (%)	97.7	91.5	84.6	71.8	75.0
false detection rate (%)	11.6	7.8	7.7	9.9	0.0

Table 3. Results based on vehicle class and type of error.

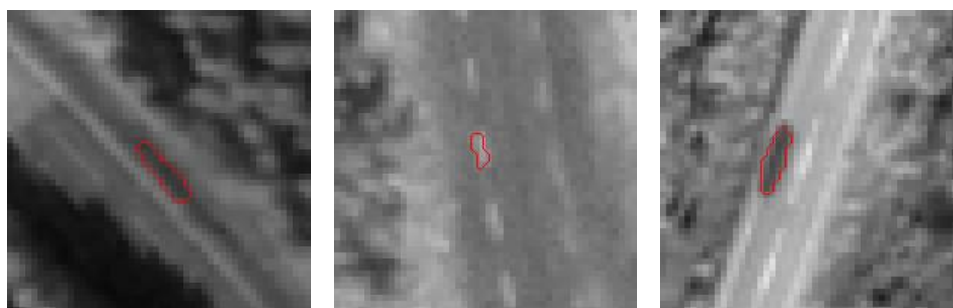|	pass. car	truck	mc	total
true vehicles (manual count)	239	68	7	314
system detected	229	60	6	295
true detections	205	57	6	268
false detection	24	3	0	27
missing detection	34	11	1	46
system detected, wrong type	4	20	6	30
false detection, double count	2	2	0	4
segmentation failed	11	5	0	16

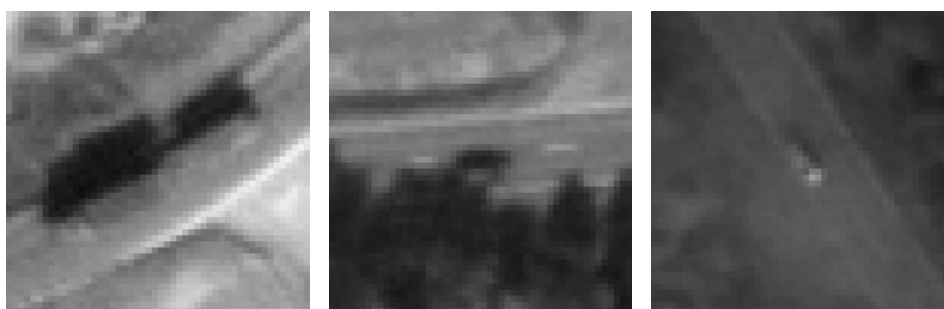Figure 5: Example false detections, from left to right: road ditch, road mark, asphalt patch.

Figure 6: Examples of missed detections, from left to right: weak contrast between truck and road, car overlaps tree shadow, bright car in cloud shadow and haze.

4. Discussion and conclusions

An approach covering all the steps needed for a fully automatic satellite based system for extraction of traffic information has been developed. The methods have been tested on a variation of scenes, covering long stretches of rural roads, and show promising results.

The detection rate varies considerably between the images, and the performance is closely related to the image conditions. Not surprisingly, the best detection rate is seen in the Østerdalen north 2004 image (true detection rate 97.7%), which is a totally cloud-, haze- and shadow-free image, with clear and orderly conditions. In the opposite end of the scale we have the Mosjøen image (true detection rate 71.8%), which contains a lot of clouds, haze and fog, making the image grainy, much of it also combined with low contrast shadow areas.

The precision of the road mask is a crucial quantity for high vehicle detection performance. For reliable operation it is necessary that the accuracy of the road mask be about 0.5m. This is an extreme situation compared to other remote sensing applications. Future work will focus on integrating contextual information, such as traffic lanes, tunnels, bus stops, etc., which may improve the accuracy of the road mask. So far, the system is developed for roads with one lane in each direction only. Since the available road vector did not include information about the number of lanes, we constructed a mask indicating parts where the road in the image has more than one lane in each direction (this applies only to minor parts of the images). This manual step can easily be excluded when the number of road lanes is included in the road vector data. However, future work should also extend the algorithms such that we are able to handle an arbitrary number of lanes.

The method would need generalization if it were to be used for other types of highway environments than those mentioned. We are currently planning a test phase where the presented methods will be validated on a large selection of images.

Acknowledgements

References

[1] Leitloff J, S Hinz & U Stilla, 2010. *Vehicle detection in very high resolution satellite images of city areas.* IEEE Transactions on Geoscience and Remote Sensing, Vol. 48, No. 7, July 2010.

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

[3] Jin X & C H Davis, 2007. *Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks.* Image and Vision Computing, Vol. 25, No. 9, pp. 1422-1431.

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

[5] Sharma G, C J Merry, P Goel & M McCord, 2006. *Vehicle detection in 1-m resolution satellite and airborne imagery.* International Journal of Remote Sensing, Vol. 27, No. 4, pp. 779-797.

[6] Agouris P, S Gyftakis & A Stefanidis, 2001. *Dynamic node distribution in adaptive snakes for road extraction.* In Vision Interface, June 2001, Ottawa, Canada, pp. 134 – 140.

[7] Theodoridis S, Koutroumbas K, 1999. *Pattern Recognition.* Academic Press, San Diego, California, pp. 309-312.

[8] Buzzone L & D F Prieto, 2001. *Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images.* IEEE Transactions of Geoscience and Remote Sensing, Vol. 39, No. 2, pp. 456-460.

[9] Quiñonero-Candela J, M Sugiyama, A Schwaighofer & N D Lawrence (Eds.), 2009. *Dataset shift in machine learning*. Cambridge, Mass., MIT Press.

[10] Salberg, A B, 2011. Retraining maximum likelihood classifiers using a low-rank model. Proceedings of IEEE International Geoscience and Remote Sensing Symposium, Vancouver, Canada, pp. 166-169.

[11] Le Hégarat-Mascle, S & C André, 2009. *Use of Markov random fields for automatic cloud/shadow detection on high resolution optical images*. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 64, pp. 351-366.

[12] Blostein D & N Ahuja, 1989. *A multiscale region detector*. Computer Vision, Graphics and Image Processing, Vol. 45, pp. 22–41.

[13] Larsen S Ø & A B Salberg, 2011. *Automatic vehicle counts from QuickBird images*. Proceedings of 2011 Joint Urban Remote Sensing Event (JURSE).

[14] Aldrin, M, 1998. *Traffic volume estimation from short-period traffic counts*. Traffic Engineering + Control, Vol. 39, pp. 656-660.

[15] Larsen, S Ø, M Aldrin & O Haug, 2008. *Estimating Annual Average Daily Traffic (AADT) based on extremely sparse traffic counts - A study of the feasibility of using satellite data for AADT estimation*. NR note no. SAMBA/49/08. publ.nr.no/4877.