

VALIDATION OF A SEMI-AUTOMATED CLASSIFICATION APPROACH FOR URBAN GREEN STRUCTURE

Øivind Due Trier^{a,*} and Einar Lieng^b

^a Norwegian Computing Center, Gaustadalléen 23, P.O. Box 114 Blindern, NO-0314 Oslo, Norway – trier@nr.no

^b Asplan Viak AS, P.O. Box 701, NO-4808 Arendal, Norway – einar.lieng@asplanviak.no

KEY WORDS: Segmentation, classification, multispectral Quickbird imagery, urban vegetation

ABSTRACT:

Municipalities in Norway need to develop an urban green structure plan. Traditional mapping has its limitation, since the land use is in focus and not the actual land cover. This study evaluated the appropriateness of using multispectral Quickbird images for the semi-automated mapping of green structures in urban and suburban areas. A Quickbird image of Oslo from 2 June 2008 was used. A classification algorithm was implemented in Definiens Developer. The algorithm was applied to the whole image, and tested on six randomly selected subsets. The validation was performed by manual editing of the classification result. The main focus of the editing process was to detect misclassifications between grey areas (such as roads and buildings) and green areas (trees, grass, and sparse vegetation). The most striking problem with the automated method was that the object borders were very rugged. However, these segmentation problems were to some extent ignored in the evaluation process, concentrating on correcting major parts of objects being misclassified rather than correcting all minor segmentation inaccuracies. The classification step had approximately 9% misclassification rate in the two-class problem grey area versus green area. This is a very good basis for further improvement. The obvious segmentation problems are clearly the first things to address when further improving the method. Another problem is to what extent the automated method can be used on other images with different light conditions, e.g., with the presence of clouds or light haze and another solar elevation. Will a simple retraining of the classification rules be sufficient, or will the rules have to be redesigned? It could even happen that redesigning the rules is not sufficient, so that other methods have to be developed.

1. INTRODUCTION

This work was initiated to meet the need of municipalities in Norway to develop a green structure plan. Traditional mapping has its limitation, since the land use is in focus and not the actual land cover. Therefore, other sources of information about urban and suburban green structure are being sought. A municipality is interested in a green structure plan for several reasons:

1. To map current status of green areas and their changes over time. For example, what happens with the vegetation in public parks over time, even if the mapped land use does not change?
2. To maintain biological diversity. Different species or groups of species use different varieties of green structure as corridors. For example, small birds would avoid open areas, and need a corridor of trees to move safely. In open areas, they would expose themselves to predators.
3. Green structures are being used for recreation.
4. Vegetation converts carbon dioxide to oxygen, reduces noise, and has aesthetical value. Vegetation also binds water, reducing the prospect of floods after heavy rainfall.
5. If accurate, the green structure map can be used in overlays

The green structure includes private gardens. Although not accessible to the public, private gardens containing trees contributes to items 2 and 4 above.

Forest and farmland are not in the focus of this study, since they are well mapped, and the land cover aligns well with the land use classification of traditional mapping.

The purpose of this study was to evaluate the appropriateness of using Quickbird 0.6 m – 2.4 m resolution satellite images for the automatic mapping of green structures in urban and suburban areas. The rest of the report is organized as follows: Section 2 presents the available Quickbird image data, followed by a description of the segmentation, training, classification and postprocessing steps of the automatic algorithm in Section 3. In section 4, the validation methodology is described. The validation results are presented in Section 5 and discussed in Section 6. This paper is a condensed version of a project report (Trier, 2009), available at <http://publ.nr.no>.

2. DATA

The project has acquired parts of a cloud-free Quickbird scene of parts of Oslo and surrounding area, acquired on 2 June 2008. The image has a 0.6 m ground resolution panchromatic band, and four 2.4 m resolution multispectral bands (blue, green, red and near infrared).

3. CLASSIFICATION PROCEDURE

Definiens Developer (Definiens, 2007) was used to segment the image, based on pixel colors and parameters describing the segment shapes. Then the user defined a set of rules to classify the segments based on texture, neighborhood, color and other attributes. The final classification result consists of five classes: (1) grey areas, (2) grass, (3) trees, (4) little vegetation, and (5) water and missing data.

* Corresponding author.

3.1 Segmentation

The segmentation was done in two levels in a bottom-up fashion.

The segmentation has to be a compromise between conflicting needs. On one hand, one would like to obtain large building blocks. At the same time, one would like to keep narrow corridors of green structure. Multiresolution segmentation was used, with two levels. The level 1 segmentation was based on the panchromatic image alone, whereas the level 2 segmentation also used the multispectral image bands (Table 1). The level 2 segmentation is based on the level 1 segmentation, which means it is locked to the segment boundaries that were created in level 1. The level 2 segmentation essentially aggregates segments from level 1.

Table 1. Segmentation parameters in Definiens Developer.

Level settings	level name	
	level1	level2
Level Usage		Create above
Image layer weights		
QB_PAN	1	1
QB_NIR	0	1
QB_Red	0	1
QB_Green	0	1
QB_Blue	0	1
Thematic layer usage	(not used)	(not used)
Scale parameter	20	50
Composition of homogeneity criterion		
Shape	0.1	0.1
Compactness	0.5	0.5

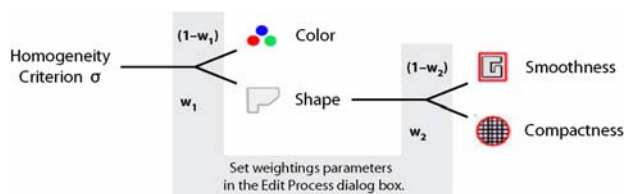


Figure 1. Homogeneity criteria in Definiens Developer. The figure is from (Definiens 2007), page 160.

On each level, the segmentation process iterates several times. In the first iteration in level one, all segments are one pixel each. The mutually best pairs according to a homogeneity criterion are found, and each identified segment pair is merged into a new segment. This continues as long as segments can be merged without breaking the scale parameter constraint. The scale parameter is a threshold on the homogeneity value of a segment, and the homogeneity value is computed as the standard deviation from the ideal situation. The following criteria can be used, in combination

- Color: homogeneity is computed as standard deviation of the spectral colors.
- Shape: divided into smoothness and compactness
 - Compactness: homogeneity is computed as the deviation from a compact object
 - Smoothness: homogeneity is computed as the deviation from a smooth object boundary.

The color and shape weights sum to 1. Within the shape criterion, the compactness and smoothness weights sum to 1

(Figure 1). So, the shape value of 0.1 in Table 1 denotes that the shape criterion has weight 10% and the color criterion 90%. By increasing the shape weight, the segmentation will be more eager to find objects which are compact and/or smooth, and less eager to find objects with low color variation.

If, for a segment, the color homogeneity is, say, 12, the smoothness homogeneity is 48 and the compactness homogeneity is 60 then the weighted homogeneity (Table 1) is

$$0.9 \times 12 + 0.1 \times 0.5 \times 48 + 0.1 \times 0.5 \times 60 = 9.2 + 2.4 + 3.0 = 14.6,$$

which is below the scale threshold for level 1, so this segment is accepted. However, if the shape homogeneity had been set to 0.5, then the weighted homogeneity had been

$$0.5 \times 12 + 0.5 \times 0.5 \times 48 + 0.5 \times 0.5 \times 60 = 6 + 12 + 15 = 33,$$

which is above the scale threshold for level 1.

In level 2, equal weight is placed on the four multispectral bands (blue, green, red and near infrared (NIR)) (Table 1). One could place a higher weight on NIR for vegetation mapping, and also reduce the weight of blue if there is haze in the image.

The scale parameter indicates how large objects one is interested in. To find individual trees, a low value should be used. To segment parts of a forest, a large value is used. We are interested in private gardens, where trees are present but the pattern is less homogeneous than in a forest. So we are interested in single trees and groups of trees, and a value of 50 seemed to work well.

3.2 Classification

The classification was done in a hierarchical fashion. At each level, there are competing rules, and the rule that gives the highest score is selected. (In the documentation, the rules are called membership functions (Definiens, 2007).) There is also a threshold for setting an object to be unclassified. This was set to 0.1. One can set this to, say, 0.9 during training.

The rules used on the 2008 Quickbird image are outlined in Figure 2, and the actual values for the thresholds should be adjusted for a new image. However, one may also want to use different rules for another image, due to different colors, phenological cycle, date, haze, etc.

Both the panchromatic 0.6 m resolution and the four bands multispectral 2.4 m resolution information was used in the classification procedure.

The classification rules are organized in a hierarchical fashion (Figure 2). Note that so-called soft thresholds are being used. This means that instead of using a simple if-test on a threshold value, essentially producing a sharp transition from 0 to 1, there is a smooth transition zone where the response goes gradually from 0 to 1. Then the rule with the highest score wins. The actual threshold values are given in (Trier, 2009).

When working with the rules, one might add new rules or tune the thresholds. At the end, one has a handful of misclassified and unclassified objects. One may then add "cleanup rules". Six cleanup rules were used, see (Trier, 2009) for details.

3.3 Comments

The segmentation and classification modules in Definiens Developer provided a means to quickly obtain a fairly good

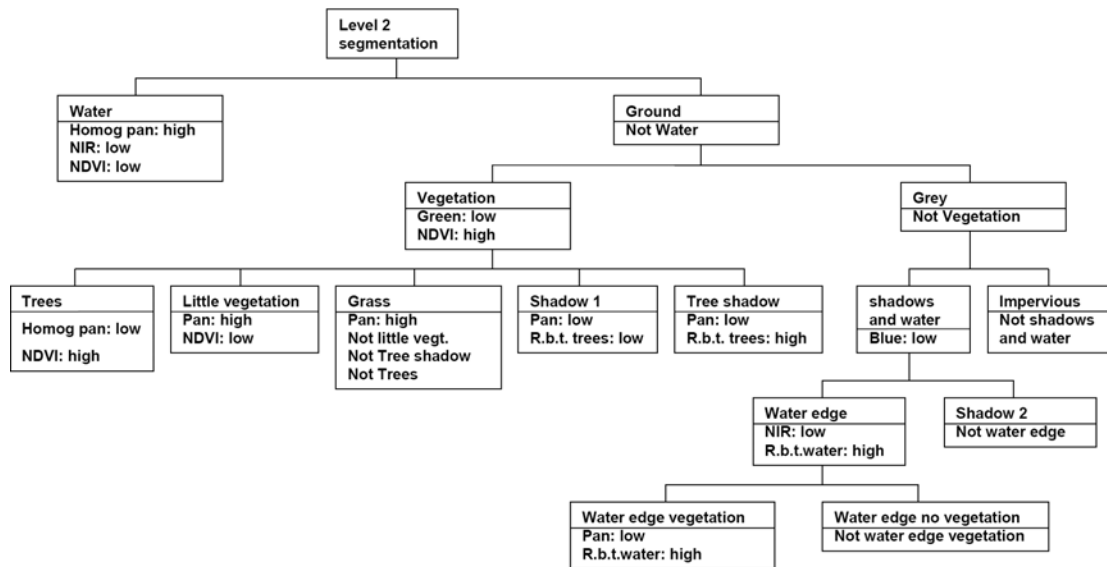


Figure 2. Hierarchy of classification rules.

classification result. Some time was spent on optimizing the parameters, but it was felt that it was not a good idea to spend too much time on this, as this would have to be repeated for a new model.

Agricultural land, rivers and lakes are not considered important in this project, as they are well mapped, and can be obtained from GIS. However, the positional accuracy is often lower than for buildings and roads.

The result of the classification procedure was a 0.6 m resolution image with the following classes:

1. Open grass land and lawns.
2. Bushes, trees, forest. (Parts of) private gardens are expected to fall into this class.
3. Little vegetation: Paths, grass areas with substantial wear and tear.
4. “Grey” areas, that is, covered by buildings, roads, parking lots, etc; thus with no vegetation.
5. Not classified or missing data, also used for water.

The three first classes are regarded as “green” areas, and can be seen as subclasses of green areas.

3.4 Postprocessing of classification result

The classified image can be combined with GIS data of buildings and roads. Trees overlapping buildings and roads are kept, based on the NDVI value, but other parts of the buildings and roads are subtracted from the vegetation classes.

Enhanced versions of the Oppegård and Lørenskog areas were created by using GIS data for buildings and roads. The houses and roads were subtracted from the green areas if the NDVI was low. In cases where the NDVI was high, for example, caused by a tree overlapping a house or a road, the tree was kept.

4. VALIDATION METHODOLOGY

The classification may be validated manually or automatically. In order to perform an automatic validation, a ground truth must be established. For Oppegård and Lørenskog municipalities, we

have obtained digital maps, free of charge, of roads and buildings, for use within the project. These maps can be used to validate grey versus green area classification, but can not be used to validate which of the three green area classes that has been assigned.

One major shortcoming of the digital map we had access to is that not all grey areas are included. Large parking lots are missing, as well as private driveways. So, the digital map could be used to find houses and public roads that were partially or fully missing in the automatic classification. However, areas that had been misclassified as grey areas could not be flagged, since many grey areas are missing in the digital map.

Thus, manual validation of the automatic classification was needed. The intention was also that the manual classification be used to validate the subclasses of green areas. However, this turned out to be too difficult to do in a quantitative manner. Only some general observations could be made. Where available, the digital map was used to guide the manual validation

4.1 Manual validation method

4.1.1 Selection of validation area

Given the size of the image, and the available resources for the project, a complete inspection of the classification result of the entire image was considered infeasible. Instead, a selection had to be made. Manual selection of areas that could be considered “representative” would lead to a biased result. On the other hand, some of the selected areas should cover the areas of which we had map coverage. These considerations led to the following selection procedure of validation areas.

1. Set the image counters $N_{Oppegård}$, $N_{Lørenskog}$ and N_{Oslo} all to zero.
2. Pick an x,y coordinate within the image at random. The range of possible values are $1 \dots x_{max}-x_{size}$ for the x coordinate, and $1 \dots y_{max}-y_{size}$ for the y coordinate, with x_{max} , y_{max} being the Quickbird image size and x_{size} , y_{size} being the validation area size.

3. If the validation area only contains missing or no data, discard the area and jump back to step 2 above.
4. If the new validation area partially overlaps an existing validation area, then replace the overlap with missing data in the new validation area
5. Compute the fraction of the area within the Oppegård map coverage ($f_{Oppegård}$), within the Lørenskog map coverage ($f_{Lørenskog}$), outside map coverage (f_{Oslo}), and with no or missing data (f_{Nodata}). These four fractions should sum to 1.
6. Add the map fractions to the counters, for example, $N_{Oppegård}(i+1) = N_{Oppegård}(i) + f_{Oppegård}(i+1)$, where i and $i+1$ denote iterations i and $i+1$, respectively.
7. Continue, by jumping back to step 2 above, until all three counters are above predefined thresholds $M_{Oppegård}$, $M_{Lørenskog}$ and M_{Oslo} .

The Quickbird image size, $(x_{max}, y_{max}) = (28090, 36602)$, and the validation area size $(x_{size}, y_{size}) = (1000, 1000)$. The validation thresholds are $M_{Oslo} = M_{Oppegård} = M_{Lørenskog} = 2$. Initially, we intended to have M_{Oslo} much higher, but the manual editing was so time-consuming that we ended up with $M_{Oslo} = 2$.

4.1.2 Validation of automatic classification

For each validation area, make a copy which is then edited, as described below. The difference between the validation area and the edited version is then used to compute a confusion matrix, counting the number and type of misclassification. Although the editing is object-based, see below, the counts in the confusion matrix are pixel-based.

For each validation area, the classified image is compared with the original image and an aerial orthophoto with 0.5 m resolution or 0.1 m resolution (Oppegård, Figure 4). All obvious misclassifications are corrected. The editing is mainly object-based, that is, individual pixels are not edited. The classified image has quite rugged object boundaries, many which could have been cleaned by using road and building outlines as a guide in the segmentation process. Noting this, we have, to some extent, avoided editing these rugged boundaries.

On some occasions, however, what should have been two or more objects have by mistake been segmented into one object only. In such cases, the object has been split and parts of it reclassified in the editing process.

On some occasions, parts of water bodies have been mistaken as grey areas, probably due to wind patterns. Since water bodies can be easily removed by using GIS data, we have not counted these as misclassifications, but regarded them as missing/no data.

Although originally intended, a validation of the three subclasses of green areas is not performed. Only a few occasional substitutions of one subclass of green with another are done.

During the manual verification, the need for a gravel subclass emerged. This class has been used in some instances to denote grey areas that are not sealed, and thus may be recovered as green areas. This is indeed the case for construction sites. Typically, when a new house is being built, the entire garden looks like a grey area in the Quickbird image, but is planted shortly after. In practice it is difficult to see the difference

between gravel, asphalt and concrete, so the gravel class is only used in very obvious occasions. It is in practice a subclass of grey areas.

5. VALIDATION RESULTS

The manual validation procedure, as described in section 4, was applied, resulting in 6 validation areas. Of these, two were from Oppegård, two from Lørenskog, and two from Oslo. The overall classification performance is about 89% correct classification rate (Table 2). This figure hides the fact that the object boundaries from the segmentation step are far from ideal. Further, in the manual validation procedure, almost no objects from one of the three green structure classes were reclassified as another green structure class. In this respect, it is more meaningful to look at the two-class problem: green versus grey areas. In this case, the recognition performance was slightly better, about 91% (Table 3)

Table 2. Classification performance when using six classes.

correct classification	89.13%
misclassification	10.87%
total	100.00%

Table 3. Classification performance when using two classes.

correct classification	91.38%
misclassification	8.62%
total	100.00%

Table 4. Combined confusion matrix for all six verification areas, in number of pixels.

Area 1-6		Edited						Sum classified
		Grass	Forest	Little vege.	Grey area	Gravel	No data	
Classified	Grass	535353	0	1	4479	110	1	539944
	Forest	931	2737263	4568	110921	2013	8650	2855696
	Little vege.	59	3164	499868	135870	3387	432	642348
	Grey area	3029	65620	178704	1575587	126162	3256	1949102
	Gravel	0	0	0	0	0	0	0
	No data	0	558	0	13	0	1	571
Sum edited		539372	2806605	683141	1826870	131672	12340	6000000

Table 5. Combined confusion matrix, in percentages.

Area 1-6		Edited					
		Grass	Forest	Little vege.	Grey area	Gravel	No data
Classified	Grass	99.25%	0.00%	0.00%	0.25%	0.08%	0.01%
	Forest	0.17%	97.53%	0.67%	6.07%	1.53%	70.10%
	Little vege.	0.01%	0.11%	73.17%	7.44%	2.57%	3.50%
	Grey area	0.56%	2.34%	26.16%	86.25%	95.82%	26.39%
	Gravel	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	No data	0.00%	0.02%	0.00%	0.00%	0.00%	0.01%
Sum edited		100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

The most common misclassification is to confuse little vegetation and grey areas. This resulted in about 300,000 pixels being reclassified (Table 4). This is about 5% of the 6,000,000 image pixels. Of the 683.141 pixels that were regarded as little vegetation after the manual validation step, 178,704, or 26%, were originally classified as grey area (Table 4 – Table 5).

6. DISCUSSION

The classification results show that the classification part of the automatic algorithm is able to classify between green and grey areas, with approximately 10% misclassification. This is clearly a good starting point for improvements. However, the

ruggedness of objects suggests that the segmentation step of the automatic algorithm has a great potential for improvement.

Another issue is to what extent the automatic algorithm can be used on another Quickbird image or not. The classification rules in the automatic classification method have been trained on a subset of the image, and then evaluated on random portions of 1000 by 1000 pixels. The illumination conditions were very close to ideal and uniform over the entire scene, whereas many other Quickbird images of Oslo have clouds. It is possible that the classification rules will have to be adjusted for every image to be processed. Also, it is not known what problems the presence of clouds will result in. All in all, it could happen that redesigning the rules is not sufficient, so that other methods have to be developed.

One minor issue was dealt with wrongly in the manual evaluation procedure. Whenever a house or road was partly obscured by a tree, the tree was ignored and the house or road was edited to show its extent. However, in the context of green structure, one is more interested in the trees than in the houses and roads. So, some correct classifications have been marked as misclassifications. However, the total number of pixels that have wrongly been edited in this manner, is small, so the main findings of the evaluation are still valid.

The smallest mapped area is approximately 100 m². If for example there is a piece of grass land in a private garden of 10 by 10 meters, then it will be mapped. However, if a medium to large tree appears in the middle, then the homogeneity criterion may flag the entire area as forest.

Private gardens appear as a mixture of the three green structure classes in addition to the houses and driveways. Gardens also contain a mix of different materials in addition to vegetation, including furniture, trampolines, etc. In the classification rules, there are additional classes. Many of these are merged into the grey area class. In addition, there are two shadow classes, one for tree shadows, which are regarded as part of green vegetation, and one for other shadows.

The manual editing resulted in an additional class: gravel, which is considered as grey area. This class was added mainly to meet a potential need to indicate temporary grey areas, and was used on construction sites. Gravel also indicates an area that is not sealed, permitting water drainage. However, gravel and sand is difficult to discriminate spectrally from concrete.

6.1 Segmentation

The results of the segmentation step are not directly available to us in the classified image, since neighboring segments in many cases have been assigned the same class in the classification step. From the classification result, it is obvious that the object boundaries of classified grey areas deviate substantially from the true outlines of houses and roads. This is especially true in suburban areas (Figure 3), where there are a lot of small roads and buildings. However, the segmentation results can be examined in Definience. This was done for a few selected areas. Level 1 segmentation often creates border segments one pixel wide and very long. These pixels are often a spectral mixing of the two neighboring regions, for example, building and vegetation, or at the edge of shadows. Many roads are also segmented into many parallel narrow and long segments. In other instances, the gradual transitions between different objects allowed segments to be merged across the true object boundaries.

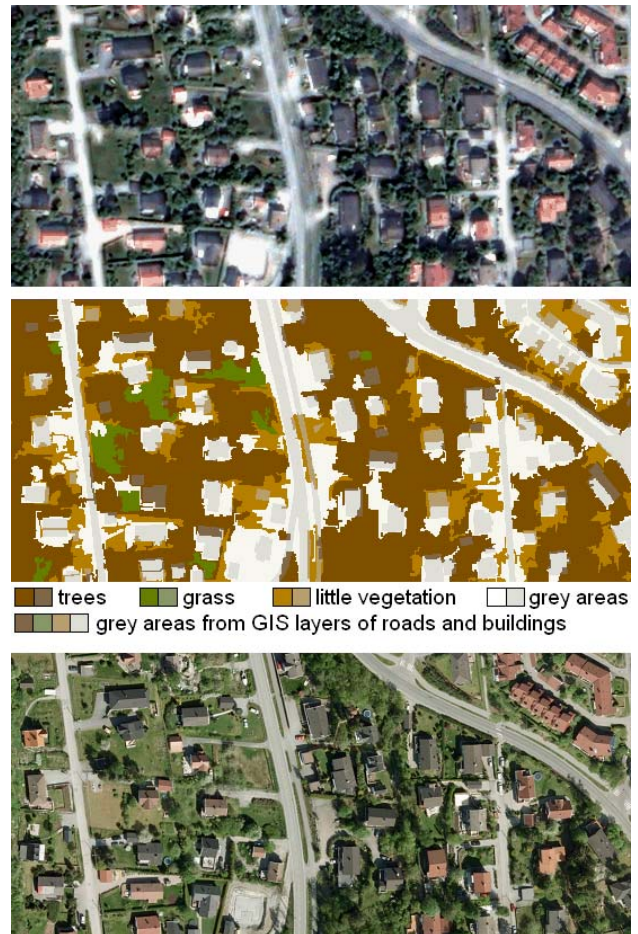


Figure 3. Segmentation problems in suburban areas in Oppegård municipality. Top: a 330 m x 250 m part inside validation area 1 of the Pan-sharpened Quickbird image. Middle: the automatic classification result for this subimage, with houses and roads from a digital map superimposed in grey. Bottom: Aerial orthophoto of the same area, captured with 10 cm ground resolution.



Figure 4. Close-up of the upper left corner of the part of the aerial image of Oppegård in Figure 3.

Many of the segmentation problems are due to shadows from buildings (Figure 5) and trees (Figure 6). Building shadows are often classified as grey areas. It could be possible to predict these shadows from the building height and the sun's position. The building height might be available from a digital map, and the sun's position can be computed from the acquisition time and date for the satellite image.

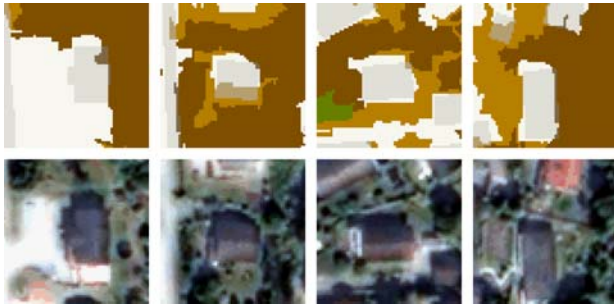


Figure 5. House shadows are sometimes misclassified as grey areas.

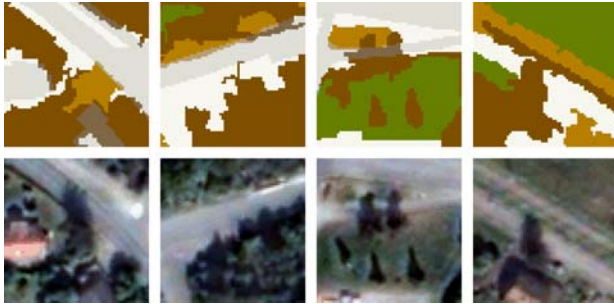


Figure 6. Tree shadows are sometimes mistaken as grey areas (far and middle left), and other times they block grey areas (far and middle right).

Tree shadows are sometimes classified as grey areas, other times they block grey areas (Figure 6). In both cases, the shadows need to be detected and removed. The tree height is not readily available, but one can make a few guesses and see if one of the heights matches the shadow length fairly well.

For both tree shadows and building shadows, the shadow outline must be extracted, and the intensity values inside the shadow increased to the level outside the shadow.

Shadows aside, there are many more segmentation issues to solve. The most important shortcoming of the current segmentation approach is that no prior information is used. By including outlines of buildings, roads, rivers and lakes from a digital map, the outlines could be used to guide the segmentation step so that the outlines from the map were preferred to some extent. In some cases, there might be coregistration errors in the order of 1–2 m between the GIS and the Quickbird image. Ideally, the segmentation algorithm should be aware of this uncertainty and allow that a, say, house be moved 1–3 pixels.

6.2 Time series of chlorophyll or NDVI

An entirely different approach than the current could be to use time series of medium or low resolution satellite images to directly measure the variation from year to year in chlorophyll, which is often estimated from the so-called *normalized difference vegetation index*, NDVI. The NDVI for a pixel (x, y) is computed from the *near infrared* (NIR) spectral band and the red (R) spectral band as

$$NDVI(x, y) = \frac{NIR(x, y) - R(x, y)}{NIR(x, y) + R(x, y)}$$

By using 250 meter resolution images from MODIS, or even 1 km resolution images from AVHRR, one obtains average values, in which a decrease in chlorophyll in one small area

may be cancelled by an increase in another small area within the same pixel. However, the general trend can be monitored, since these images are captured daily.

The Norwegian Computing Center has developed time series analysis algorithms for vegetation monitoring in other projects (Salberg, 2010; Aurdal et al., 2005). These algorithms could be modified for use on monitoring of green structure in urban and suburban areas. The time series analysis algorithm models change on three scales:

1. Daily variations due to imaging conditions
2. Phenological variation during one year
3. Changes from year to year.

During one year, the green vegetation goes through one cycle, which has nearly the same shape from one year to another, but with variations in the start and end dates of the summer season, as well as the strength of the peak of the cycle (Huseby et al., 2005). By eliminating the modeled changes on the daily, seasonal and yearly scale, one can detect statistically significant changes in individual pixels, and detect areas in which the green structure has been reduced or improved.

7. CONCLUSION

In the present work, Definiens Developer was used for segmentation and classification of a Quickbird scene from 2008. The result is validated in the present paper, and the conclusion is that this is a good starting point for further improvements of the method. The most striking problems are related to the segmentation. Object contours are often ragged, and do not follow the true boundaries of houses and roads very well. Another difficulty is shadows from buildings and trees, resulting in frequent misclassifications of whatever happens to be in the shadow areas.

REFERENCES

- Aurdal, L., Huseby, R. B., Eikvil, L., Solberg, R., Vikhamar, D., Solberg, A. H. S., 2005. Use of hidden markov models and phenology for multitemporal satellite image classification – applications to mountain vegetation classification. In *Proc. Int. Workshop Analysis Multi-Temporal Remote Sensing Images*, Biloxi, Mississippi, USA, May 16–18, 2005, pp. 220–224.
- Definiens Developer 7, User Guide, 2007. Definiens AG, Munich, Germany.
- Huseby, R. B., Aurdal, L., Eikvil, L., Solberg, R., Vikhamar, D., Solberg, A. H. S., 2005. Alignment of growth seasons from satellite data. In *Proc. Int. Workshop Analysis Multi-Temporal Remote Sensing Images*, Biloxi, Mississippi, USA, May 16–18, 2005, pp. 213–216.
- Salberg, A. B., 2010. Land cover classification of cloud-contaminated multi-temporal high-resolution images. Revised version submitted to *IEEE Transactions on Geoscience and Remote Sensing*.
- Trier, Ø. D., 2009. *Urban green structure – validation of automatic classification*. Norwegian Computing Center, Note No. SAMBA/39/09, 52 pp., <http://publ.nr.no/5159>.