Nationaal Lucht- en Ruimtevaartlaboratorium

National Aerospace Laboratory NLR

Executive summary

Multi scale object based detection and classification of roads and vehicles in high resolution optical satellite imagery

In the framework of defence and security applications, NLR has built a demonstration environment and did experiments to use object based image interpretation for the detection of roads and vehicles in single or multiple optical images. The usage of an object based approach provides the ability to include part of the operator knowledge into the software based interpretation process. The method of working for the optical road and target extraction in a single image is developed in a multi scale, multi method scheme in which three levels can be recognized. Firstly the course level, in which in an autonomous way Regions Of Interest (ROI's) are determined. Secondly the medium level, in which basic segmentation and classification is done within the ROI's, with some user interaction. Thirdly in the fine level, the classification result is refined based on the knowledge on object characteristics and contextual rules. On this level, considerable user interaction is required. For detected vehicles a form of type matching can be done depending on the image resolution. For detected road segments, cleaning and network building steps are applied. If multiple images are available, a change detection scheme is added at the medium level. The schemes have been implemented in a COTS software environment. The method has been tested on a high resolution dataset, consisting of multispectral aerial photographs and satellite images. The results show a relevant first achievement, but improvements are required. Options are more accurate correction for the 3D object geometry and more extended use of relational knowledge and probabilities.

Report no.
NLR-TP-2008-275

Author(s)
A. Oostdijk
M. van Persie
H.H.S. Noorbergen
J.W. van Rijn

Report classification
UNCLASSIFIED

Date
May 2008

Knowledge area(s)
Geomatics

Descriptor(s)
Object
Detection
Classification
Roads
Vehicles
High Resolution
Satellite Imagery

This report is based on a presentation to be held on the "Geobia 2008", Calgary, Canada, August 2008.

Multi scale object based detection and classification of roads and vehicles in high resolution optical satellite imagery

Nationaal Lucht- en Ruimtevaartlaboratorium, National Aerospace Laboratory NLR

Anthony Fokkerweg 2, 1059 CM Amsterdam,
P.O. Box 90502, 1006 BM Amsterdam, The Netherlands
Telephone +31 20 511 31 13, Fax +31 20 511 32 10, Web site: www.nlr.nl

Nationaal Lucht- en Ruimtevaartlaboratorium
National Aerospace Laboratory NLR

NLR-TP-2008-275

Multi scale object based detection and classification of roads and vehicles in high resolution optical satellite imagery

A. Oostdijk, M. van Persie, H.H.S. Noorbergen and J.W. van Rijn

This report is based on a presentation to be held on the "Geobia 2008", Calgary, Canada, August 2008.

Customer	National Aerospace Laboratory NLR
Contract number	----
Owner	National Aerospace Laboratory NLR
Division NLR	Aerospace Systems & Applications
Distribution	Unlimited
Classification of title	Unclassified
	May 2008

Approved by:

Author	Reviewer	Managing department

Contents

MULTI SCALE OBJECT BASED DETECTION AND CLASSIFICATION OF ROADS AND VEHICLES IN HIGH RESOLUTION OPTICAL SATELLITE IMAGERY

A. Oostdijk [a], M. van Persie [a], H.H.S. Noorbergen [a], J.W. van Rijn [a]

[a] National Aerospace Laboratory NLR, Space Department, PO Box 153, 8300 AD Emmeloord, the Netherlands - (arjen, persie, noorberg, rijnjw)@nlr.nl

KEY WORDS: Object based, Detection, Classification, Vehicles, Roads, High resolution, Satellite imagery

ABSTRACT:

In the framework of defence and security applications, NLR has built a demonstration environment and did experiments to use object based image interpretation for the detection of roads and vehicles in single or multiple optical images. The usage of an object based approach provides the ability to include part of the operator knowledge into the software based interpretation process. The method of working for the optical road and target extraction in a single image is developed in a multi scale, multi method scheme in which three levels can be recognized. Firstly the course level, in which in an autonomous way Regions Of Interest (ROI's) are determined. Secondly the medium level, in which basic segmentation and classification is done within the ROI's, with some user interaction. Thirdly in the fine level, the classification result is refined based on the knowledge on object characteristics and contextual rules. On this level, considerable user interaction is required. For detected vehicles a form of type matching can be done depending on the image resolution. For detected road segments, cleaning and network building steps are applied. If multiple images are available, a change detection scheme is added at the medium level. The schemes have been implemented in a COTS software environment. The method has been tested on a high resolution dataset, consisting of multispectral aerial photographs and satellite images. The results show a relevant first achievement, but improvements are required. Options are more accurate correction for the 3D object geometry and more extended use of relational knowledge and probabilities.

1. INTRODUCTION

From many applications requirements occur for fast and (semi-) automated interpretation of large high resolution optical imagery datasets. Interpretation remains an extremely complicated process however that require large involvement of the knowledge of skilled operators.

Object based approaches provide the ability to include part of the operator knowledge into the software based interpretation process. In the framework of defence and security applications NLR did an experiment to use object based image interpretation for the detection of roads and vehicles in single or multiple (multi-temporal and multi-sensor) images.

Objective of the work was to develop an integrated environment for target analyses and scene interpretation from multi-temporal high-resolution optical imagery. NLR focused on the detection and classification of roads and vehicles, the workflow and the use of change detection.

The work was related to the European Defence Agency project ERG109.035 "Satellite Ground Segment and Processing Technology" (Villmones et al., 2007). In this project NLR worked together with TNO Defence and Security (Dekker, 2008), Vexcel-NL and Imagem.

2. APPROACH

2.1 Object extraction

Automatic object extraction has been a research topic for many years, from optical images as well as from SAR images. A successful implementation not only depends on the strategy and techniques used for the extraction, but also on the type and conditions of images on which they are applied. Important factors are ground resolution, contrast with the surroundings, occluded parts, shadows, traffic, shape (straight or strongly curved), etc.

Proper pre-processing of the images before applying the object detection algorithms is important. For a successful extraction of objects it is necessary that the images are properly geo-corrected, in particular if more images are used for one analysis, either in time (time series), radiometry (multi-spectral) or in resolution (multi resolution). Also, the result will be better if certain image enhancements, like local stretching, are used before the object extraction.

2.2 Multi scale – multi method

In literature, there are many methods described to extract line and point objects, both pixel based and object based (Bacher & Mayer, 2005; Gerhardinger et al., 2005). As the goal is to extract *objects,* the object based methods are preferred to the pixel based methods. Because the classification result depends on so many variables, it is impossible to choose one method that will give the best result in all cases. Therefore, the approach should be a multi scale – multi method one (Oostdijk et al. 2007). Multi scale, because there are different resolutions involved and different sizes of objects to be extracted. Multi method, because different image contents are involved and different properties of objects have to be extracted. The approach is outlined in Figure 1.

Three levels of processing can be recognized:

- *Course level*. On this level, regions of interest (ROI's) are determined. These are the regions that are *not* classified as water, buildings etc. (all the regions where there are probably no objects). Determination of ROI's reduces the processing time in the next steps, because certain areas can be excluded from further investigation. On this level, the use of supporting data is also very important, like maps of vegetation and water.

- *Medium level*. On this level, the basic segmentation and classification is done within the ROI's. Line features that could be road segments are extracted but no mutual relations between them and the surroundings are established yet. Relatively large point objects could be detected within this level, depending of the resolution of the images and the properties of the objects.
- *Fine level*. On this level, the classification of level 2 is refined. From here, the road extraction differs significantly from the vehicle extraction scheme. Line features that could be road segments are separated from line features that are (probably) not road segments. Important inputs on this level are the road context relations and road models. At this point, the preparation for the road network generation is also done. Road segments are merged and connected and small segments that don't belong to the road (dangles) are removed. The last step on this level is the road network generation (topology). Mutual relations between the road segments and their surroundings are established and a road topology is created. This step is supported by road models and network topology rules. For the vehicle extraction, the important inputs are the object characteristics and rules, in particular the contextual rules. If the resolution is high enough, the last step is *type matching*. This method can be useful if a specific type of vehicle has to be detected, for example a car or truck. Because this method is used on the fine level, only specific areas have to be processed where already have been detections of objects. This will speed up the procedure.

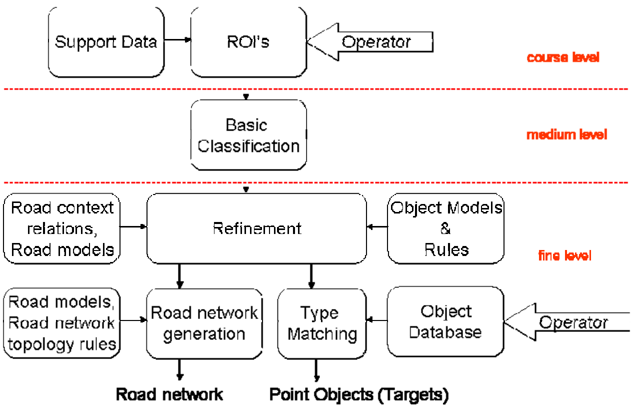

Figure 1: Object extraction strategy

2.3 Workflow

In addition to the multi scale – multi method approach, a workflow was defined according to the level of autonomy:

- *Level 1 - Define ROI's.* On this level, the ROI's are determined on the basis of (external) GIS information. This process is usually autonomous.
- *Level 2 – Object extraction.* On this level, the objects are extracted and classified in predefined object types. This process is mostly automatic, but may require some user interaction, depending on the required level of detail.
- *Level 3 – Post processing.* On this level, the result of level 2 is checked and – if necessary – adjusted. This is a *computer aided manual* process, which requires much user interaction. Also, the type matching and road network generation is done on this level.

This workflow is depicted in Figure 2.

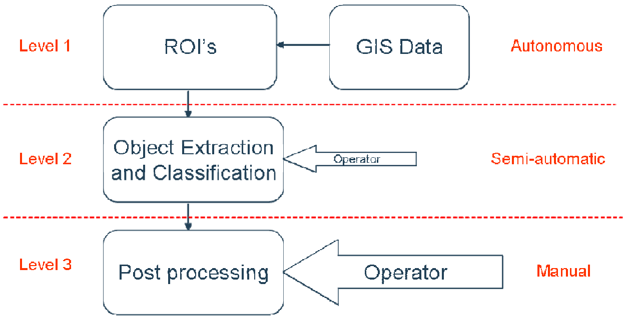

Figure 2: Workflow

3. IMPLEMENTATION

3.1 Software

For practical reasons, it was decided to use as much custom-off-the-shelf (COTS) software as possible. In this way, software development costs are minimized and the system can benefit from future software updates. Therefore, the object extraction scheme is implemented in *Definiens Professional 5.0,* a software package for object based classification, formerly known as *eCognition* (Definiens, 2005). The pre-processing is implemented in Erdas Imagine 9.1, an extensive image processing package for remote sensing data with supplementary modules for specific functionality. The creation of ROI's and the post processing are implemented in ESRI ArcGIS 9.1.

3.2 Level 1 – ROI's

On level 1, the ROI's are determined automatically on the basis of (external) GIS information, usually provided in one or more shape files. It is also possible that the user defines the ROI's him- or herself (using for example ArcGIS), but this requires of course some user interaction, see Figure 3. On this level, it is also possible to define the *objects of interest (OOI's)*, like certain kinds of vehicles or roads.


Figure 3: ROI - 500m buffer around specific locations

After the image data is pre-processed (co-registration and radiometric correction), the image data is loaded into Definiens, together with the ROI information. Using the ROI information as a mask, a very course first segmentation of the image is performed, dividing it into ROI's and non-ROI's. This process is autonomous. Also, if OOI's are defined, this information is used to define the classes for the classification in level 2.

3.3 Level 2 – Object extraction and classification

3.3.1 Single image

As test data for a single image object extraction, Quickbird images and aerial photos of the harbour area Maasvlakte in The Netherlands were used, see Figure 4.

Figure 4: Single image test data

In level 2, the pixel based image is transformed to an object based image by a segmentation step. Because this only takes place within the ROI's, a lot of time can be saved. For extracting road segments and vehicles, usually a *multiresolution segmentation* is used. In contrast to chessboard segmentation and quad-tree segmentation, multiresolution segmentation delivers objects all different in size and shape. Parameters that control the segmentation are scale, shape and smoothness. At this point, the use of additional support data, like vegetation and water maps, can improve the results.

In preparation of the basic classification, a second segmentation could be done, depending on the scale of the objects that have to be extracted in the next step, see Figure 5.

Figure 5: Multiresolution segmentation (l) and spectral difference segmentation (r)

Next, the basic object classification is done. Parameters that are taken into account are – amongst others - spectral smoothness, length, width and length/width ratio. This results in a classification, in which linear or point targets are detected, but not identified yet.

In the following step, the basic classification is refined, resulting in a more or less accurate identification of the extracted objects. For this, several methods are used:

- Refine the segmentation in Definiens Professional and use the contextual classification rules. For example, if a line segment belongs to a road, there are probably more line segments in the neighbourhood. This can be checked using the contextual parameters *existence of*, *distance to* or *relative border to*. In this way, roads, railways and waterways, or vehicles, ships and planes can be separated. Also – to some extent – the type can be determined, for example highway and country road, or car and truck.
- Use additional support data. For example, bridges will be found in the water and (old) roadmaps can be used to verify road extraction. Vehicles mostly will be on or in the neighbourhood of roads.
- For roads the "merge class" functionality in Definiens is used to connect separate line segments. Also, small loose segments that are probably not road segments can be deleted using the shape and size parameters in Definiens.

In Figure 6, an example is given of a refinement of ships-classification using support data (water).

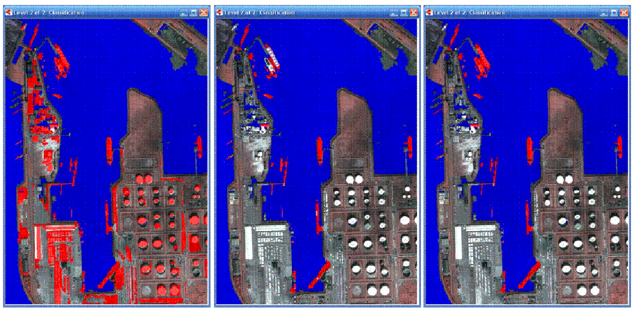

Figure 6: Basic object classification (a), cleaning (b) and refinement (c)

The result of the classification in Definiens is a shape file with features and their attributes. For the detected road segments, the topology of the line features is built, in order to generate a road network. In Figure 7, the road classification result is outlined, with the road topology and attribute table.

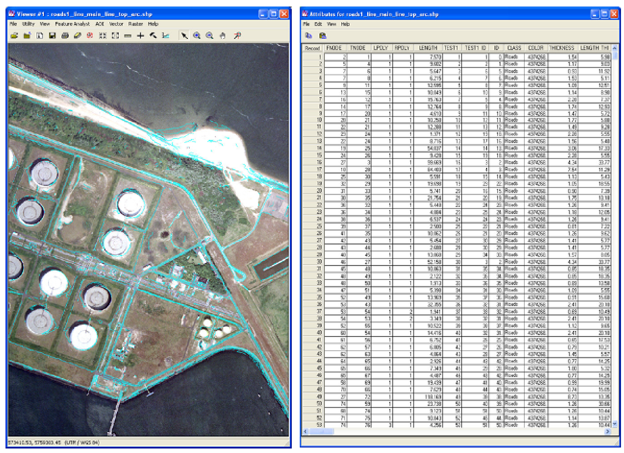

Figure 7: Road extraction result with road topology and attribute table(r)

For the detected vehicles a last type matching step is applied. In an object database the (numerical) characteristics of different types of vehicles are stored, like length, width, area, colour, etc. The characteristics of the classified vehicles can be compared to the database with SQL queries, for example in MS Access, and tie them to a specific vehicle type. The advantage of this method is that the database can easily be extended or updated. Figure 8 gives an example of a vehicle classification with the use of a very simple vehicle characteristics database, in which only the minimum and maximum length and width of cars and trucks are included. Cars are blue, trucks are purple and the red objects cannot be typed as car or truck.

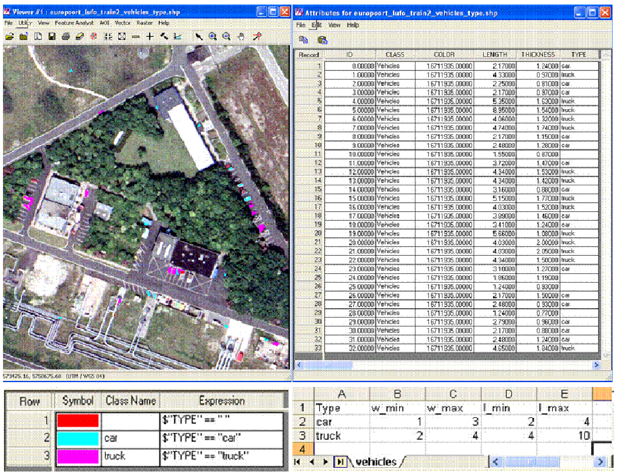

Figure 8: Vehicle classification (cars and trucks) with attribute table

3.3.2 Multiple images

The object extraction result can be improved by using multiple images, both from one sensor (multi-temporal) and from different sensors (multi-sensor).

In a next section attention is paid to fused extraction of vehicle features. Three sources of feature information are distinguished. Optical multi-temporal images, optical third party change detection results (obtained with Leica Erdas DeltaCue software) and SAR change detection results obtained from the TNO SAR change detection module (Oostdijk et al., 2007). In Figure 9, an overview is given of the object fusion process.

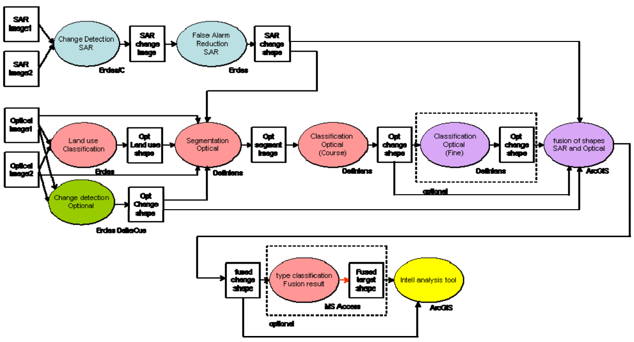

Figure 9: Fused SAR and Optical target detection, overview

The core process is the segmentation/classification of Definiens Professional. Support information, like land use classification maps, can be added at the input of the segmentation-/classification process. The three sources can be either added at the input of the segmentation-/classification process or fused at the output of the segmentation-/classification process. The last step, which is optional, is the type matching after which the output result goes into an Intelligence analysis tool, like ArcGIS, see 3.4.

Because the fusion can take place *before* as well as *after* the segmentation/classification process, a number of fusion variants are possible. Also, the change detection process within the Definiens segmentation/classification process can be implemented in two ways, see Figure 10 and Figure 11.

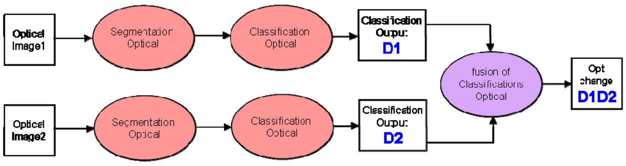

Figure 10: Definiens change detection, variant 1

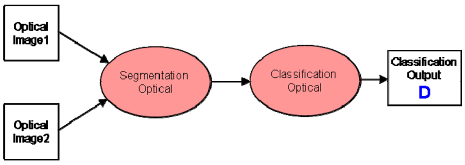

Figure 11: Definiens change detection, variant 2

All the variants were tested on a test site of Borculo in The Netherlands for which multi-temporal optical and SAR images were available and ground truth on displaced vehicles was obtained (Pride et al., 2004). In Figure 12, one of the images from the Borculo winter campaign and two enlarged subsections with the corresponding (agricultural) vehicles are outlined.

Figure 12: Daedalus winter campaign image T2

In Figure 13, the basic change detection results D, S and E are given (based on a single sensor and a single method) for both subsections where objects were placed. As can be clearly seen, the objects in the open field are detected in all three cases. The Definiens change detection (D) has the most false alarms, due to the fact that this software is not specially developed for change detection. The SAR change detection (S) detected the objects hidden under the trees because the SAR microwaves penetrate the leafless tree crowns. The DeltaCue change detection (E) is the 'cleanest' detection; it has the least false alarms.

When the results of multiple detection sources are fused, overlapping of detections takes place. The reliability of the detections is then characterised by an extra numerical attribute in the attribute table of the shape file, named *confidence*. At the moment, the confidence is directly proportional to the used number of sources contributing to the detected change. For more accurate confidence measure, it would be better to use a more complex way of calculation, taking into account the application or the sensor for example. This will be considered in future work on this subject.

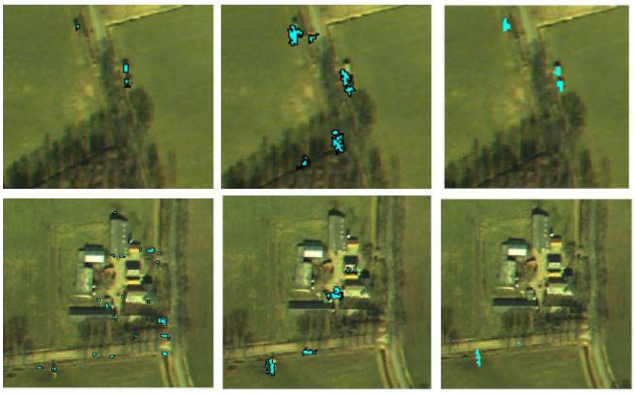

Figure 13: Basic change detection results. From left to right D, S and E

From the test results, it is clear that the more data sources are fused, the better the results are, which was to be expected. Which fusion option to choose is not only a matter of availability of software and sensors but also what is the most important property of the result (accuracy, confidence, number of false alarms) that the user is interested in. If accuracy and confidence are the most important, it is best to fuse *at the output of* the Definiens process as much as possible. The information from the separate change detections is best used in this way and does not run the risk to disappear in the Definiens segmentation process. If the (low) number of false alarms is the most important, then it is best to fuse *at the input of* the Definiens process as much as possible, because the Definiens segmentation is then driven by the SAR and DeltaCue results which will reduce the number of false alarms.

For the best *overall* results, the middle course should be adopted, for example the result DE⊕S, where the optical change detection result is added *before* the Definiens segmentation and the SAR change detection result is fused *after* the Definiens segmentation. In this result, all the objects in the open field are detected with high confidence and the number of false alarms is relatively low. Because the SAR change detection is fused at the output of the Definiens process, the objects hidden under the trees are also detected. In Figure 14, the result DE⊕S is outlined based on confidence.

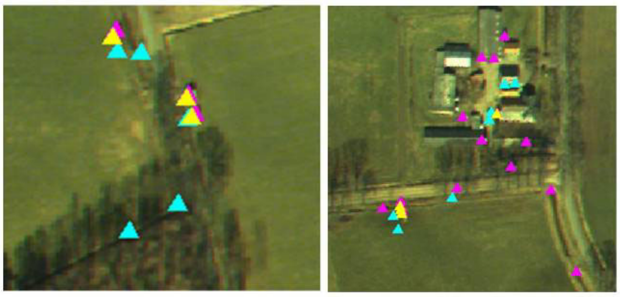

Figure 14: Daedalus winter campaign DE⊕S, based on the confidence. Cyan=1, Pink=2, yellow=3

3.4 Level 3 – post processing

In level 3, the results of level 2 are manually reviewed and, if necessary, updated. For this, an edit facility is available in ArcGIS, see Figure 15. For the roads, bad connections can be repaired and wrongly classified roads or road components can be deleted. Of course, the road network has to be rebuilt if the result has been edited. For the vehicles, wrongly classified types can be relabeled or deleted, where it is also possible to label a vehicle as "unknown", see Figure 16.

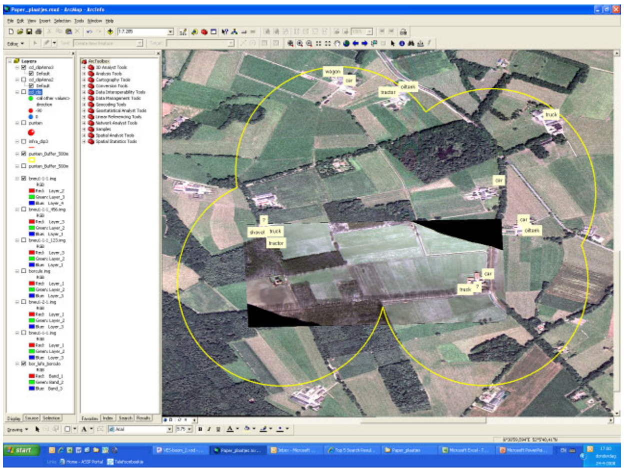

Figure 15: Review and edit facility in ArcGIS, overview

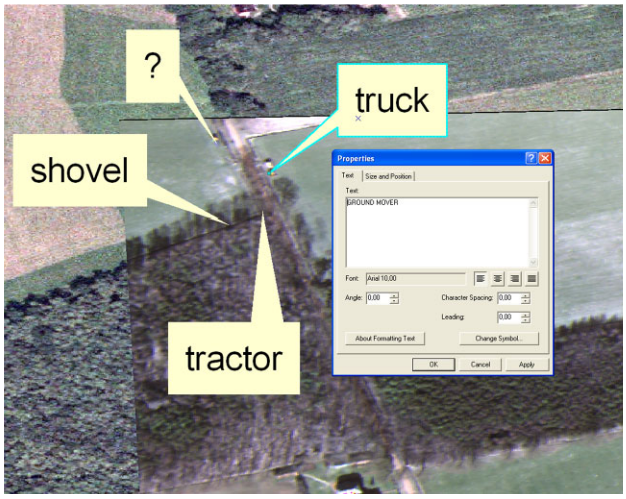

Figure 16: Review and edit facility in ArcGIS, detail

4. DISCUSSION AND CONCLUSIONS

The results of object extraction based on segmentation and classification depend heavily on the quality and resolution of the image data. Higher resolution is not always easier to classify, because more noise is introduced and the segmentation of the image data is harder and takes more time.

For road extraction, bare ground gives problems because the spectral characteristics resemble the roads very much. If the Definiens parameters can be tuned on a specific area and support data is available, results are optimal. If the process has to be automated however - which is of course preferred in this project - the parameters have to hold for all the areas as good as possible, so it will become a compromise. The use of multiple data sets, like aerial photos, Ikonos data and Daedalus data, can

improve the results, provided that the data sets are coregistered well. However, care must be taken because certain artefacts in the image data, like mosaic sewing lines and haze, can give miss-classifications.
Vehicle extraction is particularly difficult if the same segmentation parameters have to be used for rural and urban areas at the same time, because of scale differences. In rural areas, the vehicles are relatively small compared to the surrounding structures, which can make the segmentation process difficult and time consuming.

From the results presented in this article it can be concluded that the object based approach for the detection and classification of vehicles and road networks from single or multiple high resolution satellite images gives interesting first results. It provides possibilities to bring in additional knowledge on the image objects of interest, on object inter-relations and relations to external map or GIS information. It also can be used as method for fused interpretation of images from multiple sources and/or times.

It also can be concluded that the obtained results represent only a first step of the required final product. Further extension of the method is required. One element in this is to make the parameter settings dependant on image content (for this the new Developer version of Definiens provides more possibilities). Another element is to bring in additional knowledge rules in the method. A third step is to do more extended testing, resulting in more quantitative analyses of the results.

The usage of higher spatial resolution imagery has the advantage of giving more detailed information on the object (form, texture). On the other hand it introduces additional noisy features and may complicate objects (e.g. a car is not a single object, but consists of a front side, backside, roof and two windows). Another aspect is that the 3D geometry of the objects becomes more relevant, not only as object feature, but also as disturbing aspect in the sense of geometric distortion. Accurate 3D modelling and correction is an essential step therefore, that should be integrated in the methodology.

Finally, some remarks about Definiens Professional. The segmentation and classification process depend on many parameters and can become very complicated. Moreover, choices of parameters – in particular the segmentation parameters - depend heavily on the scale and contrast ratio of the images and are therefore not very generic. In the test cases described here, the complexity and number of parameters are kept deliberately low and the same for all variants, in order to keep the results reproducible and comparable to each other. In an operational environment, it would be best to adapt the parameters to the content, quality and origin of the images and group them according to the type of application. In this respect, much is to be expected from Definiens Developer, the successor of Definiens Professional, which can adapt the segmentation and classification parameters according to the image content.

5. ACKNOWLEDGMENT

The authors acknowledge the European Defence Agency and Dutch MOD for supporting the ERG109.035 project in which framework these activities were carried out, and also the Dutch partners TNO Defence&Security, Microsoft-Vexcel and Imagem who contributed to the work done in this project.

REFERENCES

Bacher U & Mayer H, 2005. CMRT05 Automatic Road Extraction from Multispectral High Resolution Satellite Images. Object Extraction for 3D City Models, Road Databases, and Traffic Monitoring – Concepts, Algorithms, and Evaluation), *Joint Workshop of ISPRS and DAGM, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Vol. XXXVI, part 3/W24, pp. 29-34

Definiens Professional/eCognition product description, 2005. www.definiens.com

Dekker R J, 2008. Road extraction and network building from synthetic aperture radar images using a-priori information. Proc. IEEE International Geoscience & Remote Sensing Symposium (IGARSS), 6-11 July 2008, Boston, Massachusetts, U.S.A.

Gerhardinger A, Ehrlich D & Pesaresi M, 2005. Vehicles Detection from Very High Resolution Satellite Imagery. *CMRT05. IAPRS, Vol XXXVI,* Part 3/W24

Oostdijk A, Dekker R J, Bozelie W C G & Knoops J, 2007. Automated Feature Extraction and Target Detection from Multi-Sensor, Multi-Data, High Resolution Images using GIS, Final Report. RES3310.8 NLR Report NLR-CR-2007-617. Amsterdam:NLR

Pride, R et.al, 2004. EU Project PRESENSE – Publishable Final Report, EU contract No. ENK6-CT2001-00553, Advantica, UK.

Villmones N J et al, 2007. ERG 109.035 Satellite Ground Segment and Processing Technology, Final Report, RES 0100.2 TromsØ, Kongsberg Spacetec AS