Nationaal Lucht- en Ruimtevaartlaboratorium

National Aerospace Laboratory NLR

Executive summary

Object based detection and classification of roads and vehicles in high resolution optical satellite imagery

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.

The method of working 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 regions of interest (ROI's) are determined. Secondly the medium level in which basic segmentation and classification is done within the ROI's. Thirdly in the fine level where the classification results are refined, based on the knowledge on object characteristics and contextual rules. In a second activity attention was paid to the fusion of change detection results of optical and SAR multi-temporal algorithms.

The results show that the object based approach 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 or times. Further extension of the method is required however, especially focusing more accurate correction for the 3D object geometry and more extended use of relational knowledge and probabilities.

Report no.
NLR-TP-2008-274

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

Report classification
UNCLASSIFIED

Date
June 2008

Knowledge area(s)
Geomatics

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

This report is based on a presentation held at the EARSeL Joint Workshop "Remote Sensing - New Challenges of High Resolution", Bochum, Germany, March 2008.

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-2008-274

Object based detection and classification of roads and vehicles in high resolution optical satellite imagery

A. Oostdijk, M. van Persie and H.H.S. Noorbergen

This report is based on a presentation held at the EARSeL Joint Workshop "Remote Sensing - New Challenges of High Resolution", Bochum, Germany, March 2008.

The contents of this report may be cited on condition that full credit is given to NLR and the authors.

This publication has been refereed by the Advisory Committee AEROSPACE SYSTEMS & APPLICATIONS.

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

Approved by:

Author	Reviewer	Managing department
18/12/8	18/12	18.12.8

Contents

Abstract	3
Introduction	3
Object detection	4
Object fusion	9
Conclusions	13
Acknowledgements	13
References	13

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

Arjen Oostdijk, Mark van Persie and Hein Noorbergen

National Aerospace Laboratory NLR, Space Department, PO Box 153, 8300 AD Emmeloord, the Netherlands; arjen@nlr.nl; persie@nlr.nl; noorberg@nlr.nl

ABSTRACT

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.
The method of working 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 regions of interest (ROI's) are determined. Secondly the medium level in which basic segmentation and classification is done within the ROI's. Thirdly in the fine level where the classification results are refined, based on the knowledge on object characteristics and contextual rules. In a second activity attention was paid to the fusion of change detection results of optical and SAR multi-temporal algorithms.
The results show that the object based approach 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 or times. Further extension of the method is required however, especially focusing more accurate correction for the 3D object geometry and more extended use of relational knowledge and probabilities.

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 defense 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 and SAR imagery. NLR focused on the detection and classification of roads and vehicles in high resolution optical imagery and on the fusion of the detected vehicles with vehicle detection results obtained by TNO with SAR images.

The work was related to the European Defense Agency project ERG109.035 “Satellite Ground Segment and Processing Technology” ([i]). In this project NLR worked together with TNO Defense and Security, Vexcel-NL and Imagem.

OBJECT DETECTION

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 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.

Multi scale - multi method approach

In literature, there are many methods described to extract line and point objects, both pixel based and object based ([ii,iii]). 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 ([iv]). 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*.

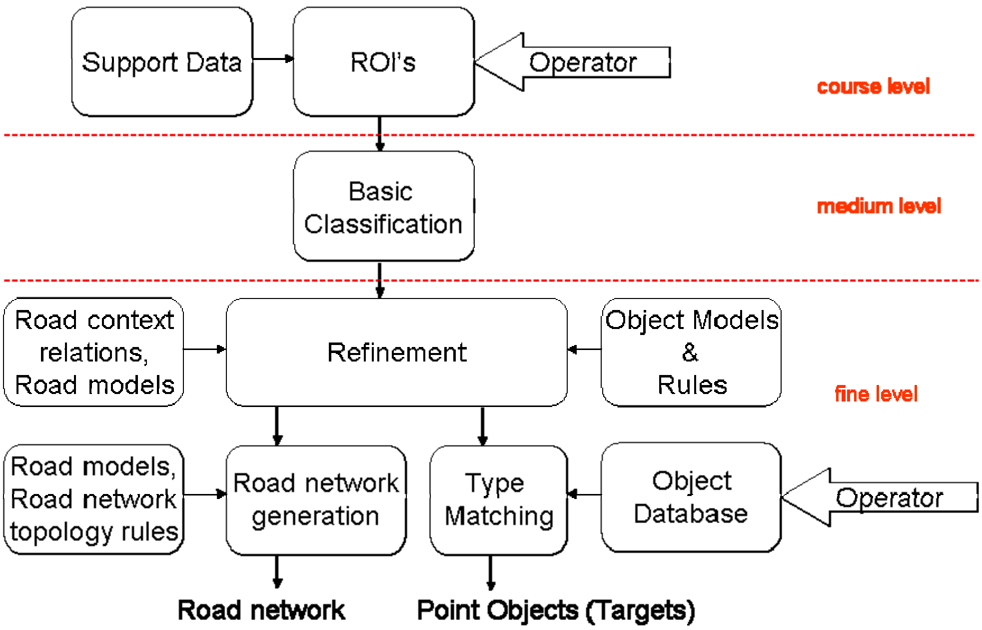

Figure 1: Object extraction strategy

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.

Implementation and results

As test data, Quickbird images and aerial photos of the harbor area Maasvlakte in The Netherlands were used, see Figure 2.

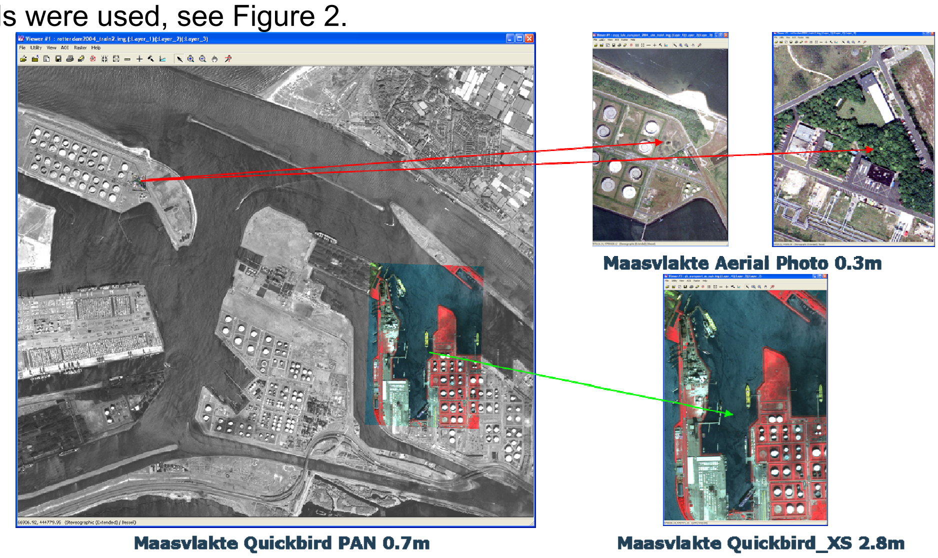

Figure 2: Object detection test data

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* ([v]). The road network generation is implemented in Erdas Imagine 9.0, an extensive image processing package for remote sensing data with supplementary modules for specific functionality.

Coarse level: After the image data is pre-processed (co-registration and radiometric correction), the image data is loaded in Definiens. The pixel based image is transformed to an object based image by a segmentation step. 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. After segmentation, regions of interest (actually the regions of non-interest) are determined with a simple spectral classification. In this way, regions of non-interest can be omitted in further processing on the one hand, and the classification in the regions of interest will be more accurate on the other hand. At this point, the use of support data, like vegetation and water maps, can improve the results.

Medium level: In preparation of the basic classification, a second segmentation on a different level could be done, depending on the scale of the objects that have to be extracted in the next step. Next, the basic linear 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.

Fine level: 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 neighborhood. 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 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 neighborhood 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. This can be a cumbersome process, that is very hard to automate, yet it is important to do before the topology is build.

In Figure 3, an example is given of a refinement of ships-classification using support data (water).

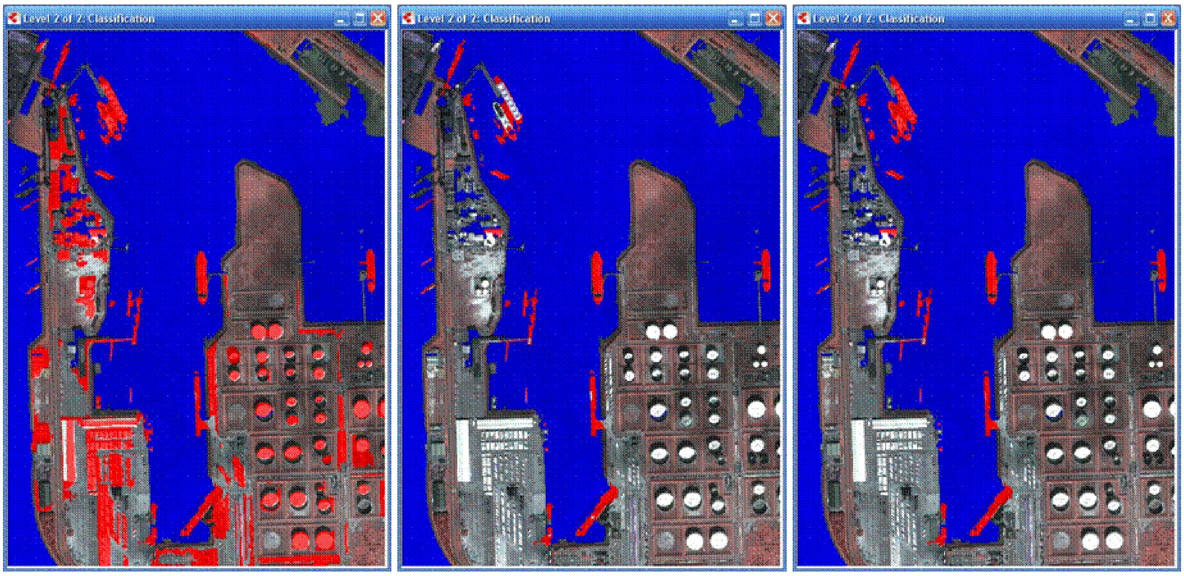

Figure 3: Basic object classification (a), cleaning (b) and refinement (c)

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 build with the vector module of Erdas Imagine 9, in order to generate a road network. In Figure 4, the road classification result is outlined, before and after refinement, with the road topology and attribute table.

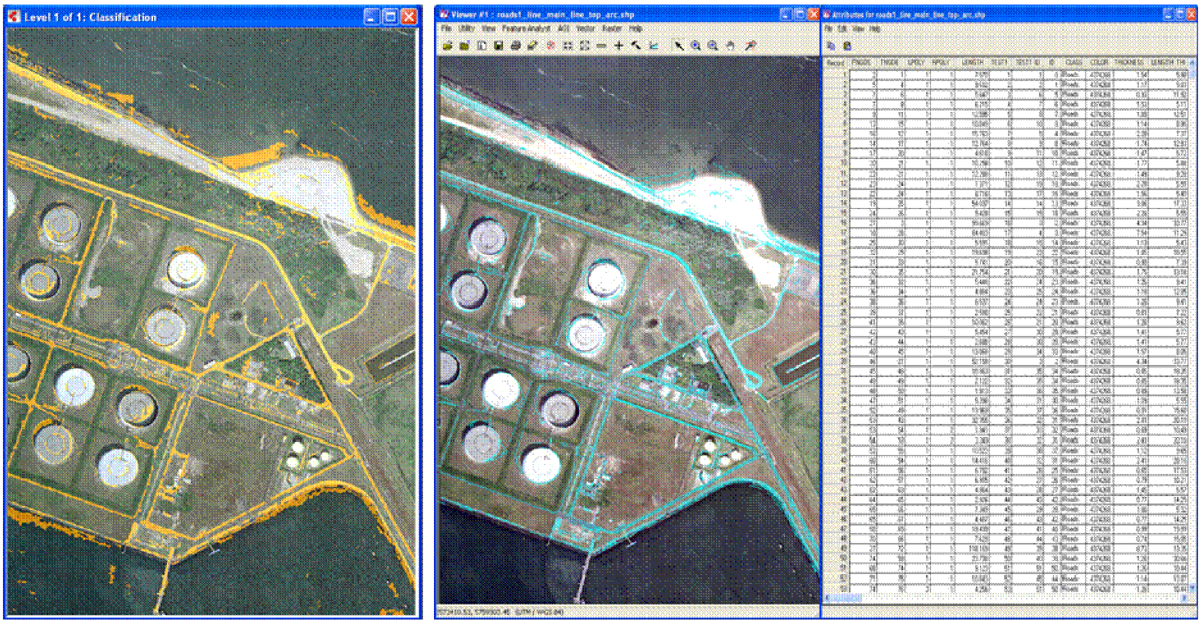

Figure 4: Road extraction before (l) and after refinement with road topology and attribute table(r)

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, color, 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 5 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.

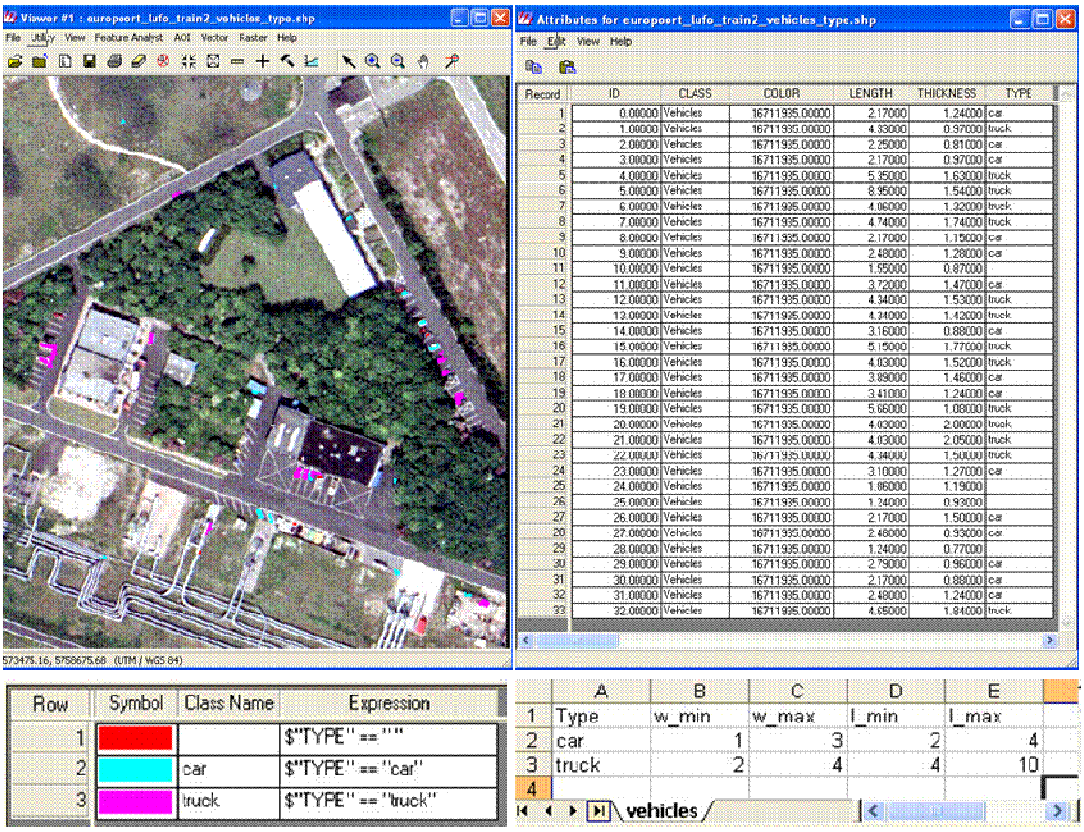

Figure 5: Vehicle classification (cars and trucks) with attribute table

The results of road extraction based on classification depend heavily on the quality and resolution of the image data. Bare ground gives problems because it is hard to classify as region of non-interest (RONI). The reason for that is that the spectral characteristics resemble the roads very much. 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. 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 coregistrated well. However, care must be taken because certain artifacts in the image data, like mosaic sewing lines and haze, can give miss-classifications.

The results of vehicle extraction based on classification depend – as with road extraction - 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. Vehicle extraction is particularly difficult if the same parameters have to be used for rural and agricultural areas at the same time, because of scale differences. In agricultural areas, the vehicles are relatively smaller than in rural areas, and this makes the segmentation process difficult and very time consuming.

OBJECT FUSION

In a next section attention is paid to fused extraction of vehicle features. Fusion focussed on the combination of the above described results with results from a SAR based detection (developed in the same project) and results of change detection from multi-temporal optical imagery. 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.

Method

In Figure 6, an overview is given of the object fusion process.

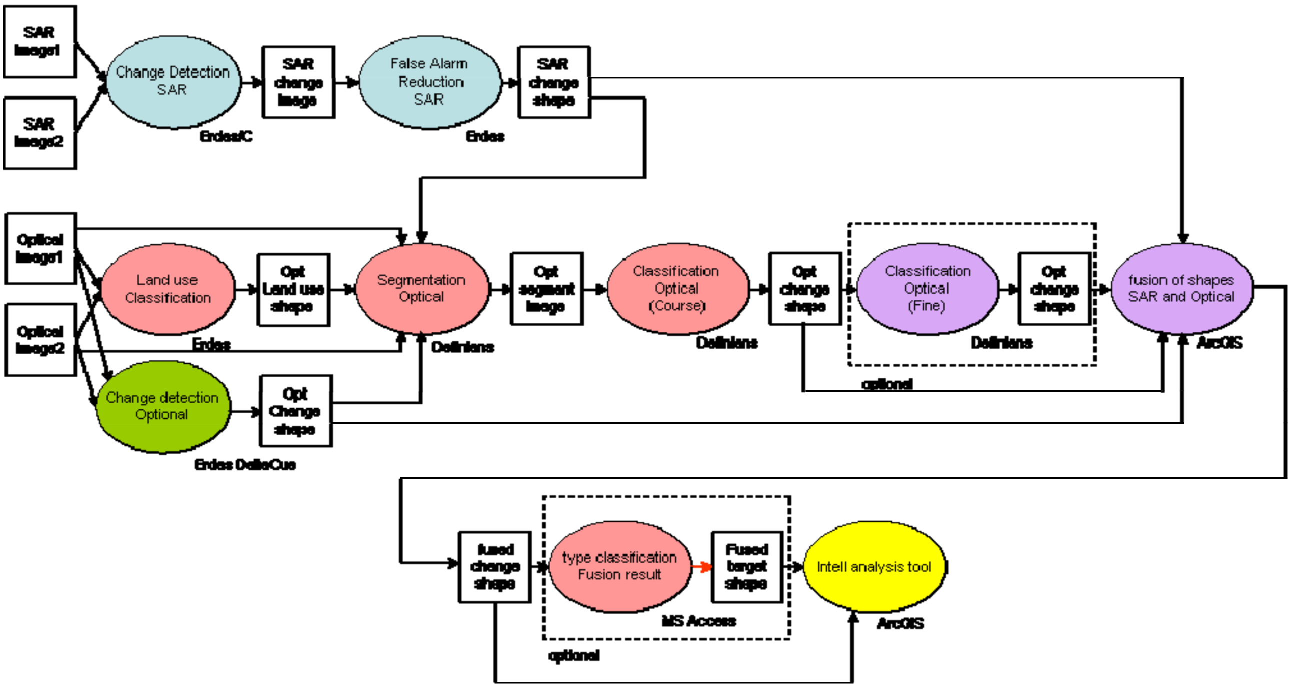

Figure 6: Fused SAR and Optical target detection, overview

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.

Implementation and results

Because the fusion can take place *before* as well as *after* the segmentation/classification process, a number of fusion variants are possible. If the Definiens segmentation/classification result is designated by D, the Erdas DeltaCue change detection result by E and the SAR change detection result by S, the variant DS⊕E means for example, that the fusion with the SAR change detection result has taken place *before* the Definiens segmentation/classification process, and the fusion with the Erdas DeltaCue change detection result *after* it.

All the variants are 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. In

Figure 7, one of the images from the Borculo winter campaign and two enlarged subsections with the corresponding (agricultural) vehicles is outlined.

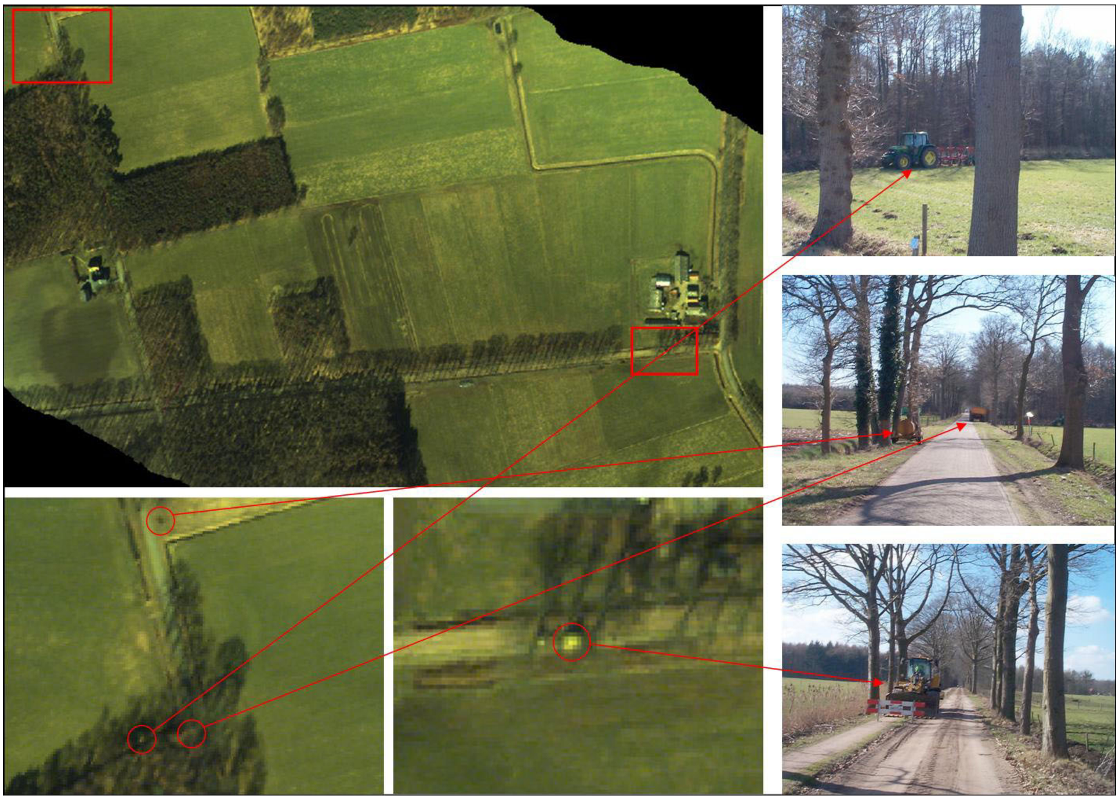

Figure 7: Daedalus winter campaign image T2

In Figure 8, 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.

Figure 8: Basic change detection results. From left to right D, S and E

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.
From the test results, it is clear that the more data sources are fused, the better the results are, which was to be expected. That means that the results D⊕S⊕E, DE⊕S and DS⊕E are the best in terms of accuracy, confidence and number of false alarms, see Figure 9. 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.

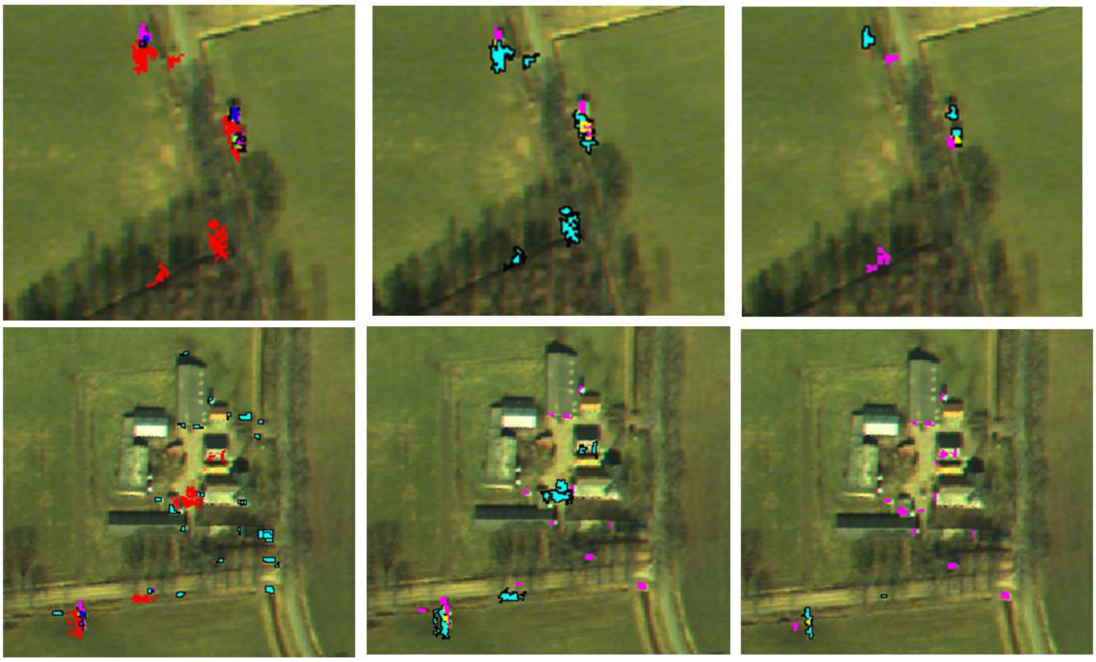

Figure 9: Fused change detection results. From left to right D⊕S⊕E, DE⊕S and DS⊕E

For the best *overall* result, the middle course should be adopted and in this case, this is the result DE⊕S. 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 10, the result DE⊕S is outlined based on confidence.

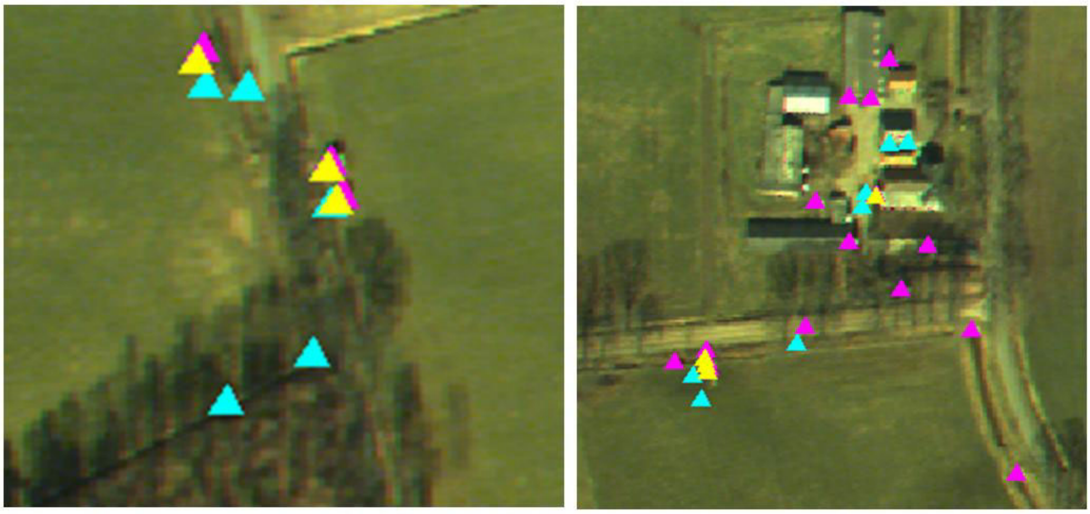

Figure 10: Daedalus winter campaign DE⊕S, based on the confidence. Cyan=1, Pink=2, yellow=3

Finally, some remarks about Definiens Professional. The segmentation and classification process depends 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 case 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.

CONCLUSIONS

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 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). At the other side 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). A third aspect is that the 3D geometry of the objects become more relevant, both as object feature, but also as disturbing aspect the in the sense of geometric distortion. Accurate 3D modeling and correction is an essential step therefore that should be integrated in the methodology.

ACKNOWLEDGEMENTS

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 en Imagem who contributed to the work done in this project.

REFERENCES

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

[ii] 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.

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

[iv] 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

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