Detecting buildings and roads from IKONOS data using additional elevation information

ABSTRACT

Using high resolution imagery such as IKONOS data should make it possible to detect man-made-features such as buildings and roads more easily than with conventional satellite image data. However, due to the higher spatial resolution of IKONOS data, an automatic or semi-automatic detection of such features based only on their spectral characteristics can become difficult, especially in heterogeneous areas such as dense urban areas (see Bauer, T. & Steinnocher, K. in this issue). A typical problem in urban remote sensing is the handling of shadows. Using a DEM and additional semantic information can help to detect such cases and to manage them adequately. Furthermore, when using an additional DEM, significant elevation information of questionable objects can be used to identify their shape. As eCognition is able to use an arbitrary number of channels for the image segmentation and classification, the DEM was used for the initial segmentation and for the subsequent object classification. Thereby, the influence of the DEM on the object generation can be controlled by adjusting the channels' weights. Based upon the underlying concepts of eCognition to generate and classify image objects, different strategies have been developed.

ZUSAMMENFASSUNG

Extraktion von Gebäuden und Straßen aus IKONOS-Daten mittels Höheninformationen
Verwendet man hochaufgelöste Bilddaten, wie die des IKONOS Sensors, können sog. Man-made-features, wie Gebäude und Straßen, prinzipiell leichter erfasst werden, als mit herkömmlichen, satellitengestützen Bilddaten.
Wegen der höheren Auflösung der IKONOS Daten gestaltet sich aber gerade in heterogenen Gebieten, wie z.B. in dichten urbanen Räumen, eine automatische oder halbautomatische Erfassung solcher Objekte, basierend auf deren Spektraleigenschaften, schwierig (vgl. Bauer, T. & Steinnocher, K. in diesem Heft). Ein typisches Problem innerhalb urbaner Räume ist der Umgang mit Schatten. Die Verwendung eines DHMs und zusätzliche semantische Information kann dazu beitragen, solche Fälle entsprechend zu erfassen und zu behandeln. Darüber hinaus kann aus dem DHM signifikante Höheninformation der zu erfassenden Objekte für deren Gestalterfassung verwendet werden. Da eCognition zur Bildsegmentierung und -klassifikation eine Vielzahl von Kanälen verwenden kann, wurde das DHM sowohl zur Bildsegmentierung, als auch für die anschließende Klassifikation der Segmente herangezogen. Dabei kann der Einfluss des DHMs auf die Objektgenerierung durch unterschiedliche Gewichtung der Kanäle gesteuert werden. Basierend auf den prinzipiellen Möglichkeiten, die eCognition zur Objektgenerierung und -klassifizierung anbietet, wurden entsprechend unterschiedliche Strategien entwickelt.

Data and Pre-Processing

For the analysis an orthorectified IKONOS sub-scene with a ground resolution of 1m (pan) res. 4m (multi-spectral) from Tsukuba (Japan) was used. The image was acquired on 2000–03–02 at 01:13pm local time and has a size of 3042 x 3051 pixels. From experience with other IKONOS data from urban areas we came to the conclusion that it may be useful to undertake a pan-sharpening before working with eCognition. In the present case this was done by applying a simple[1] inverse principal components transformation on the image data. It has to be mentioned that it is usually not recommended to apply such enhancement methods since they change the spectral properties of the objects and thus might influence the subsequent classification. However, when working on small objects, this technique can help to enhance the initial segmentation results as it takes important boundaries into account. This usually leads to more meaningful objects even on a large scale and consequently to improved classifi-

Dipl.-Geogr. Peter Hofmann

joined DEFiNiENS Division Imaging and Geomatics in January 2000. He is responsible for Piloting and Application Developement in the field of Remote Sensing & GIS.

Adresse: DEFiNiENS AG, Rindermarkt 7;
80331 Munich, Germany
Tel.: +49-(0)89/231180–38
Fax: +49-(0)89/231180–80
E-Mail: phofmann@definiens.com
www.definiens.com

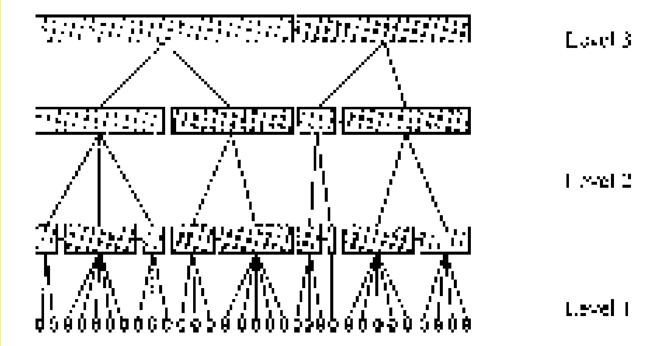

Fig. 1: Hierachical network of image objects (Source: Willhauck, G., 2000)

cation results when using dedicated shape features. Nevertheless one should keep in mind that the objects' spectral statistics are modified, which could lead to a poorer separability of the object-classes within feature space.
As DEM we derived an ArcView shape file of elevation points from an airborne laser scanning overflight with a spacing of 1.5m where each point holds height information as an attribute. The shape file was interpolated into a raster DEM, corresponding to the position and resolution of the IKONOS data set. The DEM was georeferenced to the IKONOS image using a 2nd degree polynomial. Since the DEM showed some speckle noise we applied a 3 x 3 median filter as some segmentation tests based only upon the DEM generated irregular borders and thus would have led to image objects which are difficult to handle.

Image segmentation and classification

eCognition's segmentation approach allows to generate image objects on an arbitrary number of scale-levels taking into account criteria of homogeneity in colour and shape. Thereby a hierarchical tree of image objects is generated wherein each object knows its neighbouring objects in horizontal and vertical direction (see Baatz, M. & Schäpe, A., 1999a; Baatz, M. & Schäpe, A., 1999b; de KOK, R. et al., 1999; *figure 1*).
The basic aim of the segmentation process should be to generate meaningful objects. This means that the shape of each real-world-object is represented by a corresponding image object. This shape is used to initially classify the generated image objects by their physical properties (colour, texture and form). After this physical classification, additional semantic information can be used to improve the image classification. Regarding the multi-scale behaviour of real-world-objects, it is obvious that a number of small objects can be aggregated to form a larger object constructing a semantic hierarchy *(figure 1)*. Likewise, a large object can be split into a number of smaller objects. Modelling this behaviour is a basic task of semantic nets within image analysis (see Liedtke, C.-E. et al., 1997; Hinz, S. et al., 1999) and leads to two main approaches to image analysis: A top-down and a bottom-up approach (see Bückner, J. et al., 1999; Hofmann, P. and Reinhardt, W., 2000). In eCognition this modelling can be realised performing the following steps:

- Creating a hierarchical network of image objects using multi-resolution segmentation. The higher-level image segments represent small-scale objects while the lower-level segments represent large-scale objects.
- Classifying the derived objects by their physical properties. This also means that the class names and the class hierarchy are representative with respect to two aspects: the mapped real-world and the image objects' physically measurable attributes. Using inheritance mechanisms accelerates the clas-

Tab. 1: Segmentation parameters

Level 1:		
Channel weighting:		
DEM	10.0	
IKONOS blue (pan-sharpened):	1.0	
IKONOS green (pan-sharpened):	1.0	
IKONOS red (pan-sharpened):	1.0	
IKONOS nir (pan-sharpened):	1.0	
Scale Parameter:	[illegible]	
Homogeneity Criterion:		
Color:	0.8	
Shape:	0.2	Smoothness: 0.9
		Compactness: 0.1
Level 2:		
Channel weighting:		
DEM	0.5	
IKONOS blue (pan-sharpened):	1.0	
IKONOS green (pan-sharpened):	1.0	
IKONOS red (pan-sharpened):	1.0	
IKONOS nir (pan-sharpened):	1.0	
Scale Parameter:	35	
Homogeneity Criterion:		
Color:	0.8	
Shape:	0.2	Smoothness: 0.9
		Compactness: 0.1
Level 3:		
Channel weighting:		
DEM:	0.0	
IKONOS blue (pan-sharpened):	1.0	
IKONOS green (pan-sharpened):	1.0	
IKONOS red (pan-sharpened):	1.0	
IKONOS nir (pan-sharpened):	1.0	
Scale Parameter:	40	
Homogeneity Criterion:		
Color:	0.8	
Shape:	0.2	Smoothness: 0.7
		Compactness: 0.3
Level 4:		
Channel weighting:		
DEM:	0.5	
IKONOS blue (pan-sharpened):	1.0	
IKONOS green (pan-sharpened):	1.0	
IKONOS red (pan-sharpened):	1.0	
IKONOS nir (pan-sharpened):	1.0	
Scale Parameter:	[illegible]	
Homogeneity Criterion:		
Color:	0.8	
Shape:	0.2	Smoothness: 0.7
		Compactness: 0.3
Level 5:		
Channel weighting:		
DEM	10.0	
IKONOS blue (pan-sharpened):	0.0	
IKONOS green (pan-sharpened):	0.0	
IKONOS red (pan-sharpened):	0.0	
IKONOS nir (pan-sharpened):	0.0	
Scale Parameter:	[illegible]	
Homogeneity Criterion:		
Color:	0.8	
Shape:	0.2	Smoothness: 0.9
		Compactness: 0.1

[1] I.e. using the raw channels without any histogram equalisation

sification task while making it more transparent at the same time.

- Describing the (semantic) relationships of the network objects in terms of neighbourhood relations or being a sub- or super-object. This usually leads to an improvement of the physical classification and the class hierarchy.
- Aggregating the classified objects to semantic groups which can then be used for a so called 'classification-based' segmentation. The derived contiguous segments can be exported and used in GIS. The semantic groups can also be used for further neighbourhood analyses.

These steps describe the usual process when working with eCognition. While the first two steps are mandatory, the latter two steps may be advisable depending on the user's objectives and content of the image.

Image segmentation in urban areas using DEM and spectral data

Regarding the different shapes of urban objects – especially those of buildings – and the high spatial resolution of the IKONOS sensor, a top-down approach for the image segmentation which takes these different shapes into account is very reasonable. Beginning with a coarse segmentation this means that all sub-objects will usually be generated with respect to the boundaries of their super-objects. According to the requirement 'as meaningful as possible' the shape of the generated segments should represent meaningful objects. Ideally this reflects the embedding of large-scale-objects into a wider small-scale-context. Additionally, in some cases the creation of a very finely segmented, lowermost level has been experienced as useful. This level could act as a texture-giving layer for the higher-level objects (see Knoke, R., 1999; DEFiNiENS, 2000). On the medium and lower levels the image segments should represent objects which can be semantically considered as sub-objects of their coarser super-objects.

When using only elevation information for the image segmentation, the generated objects will either be significantly higher than their neighbourhood or mostly flat with borders determined either randomly or by the natural relief. On the higher levels, the top-down-approach yields two general types of objects on each level: large, high and compact objects as well as lower or flat objects whose overall elevation value is more variable. On the lower levels, smaller and lower but not necessarily less compact objects are obtained. Applying the top-down-approach raises the level of detail with decreasing object size.

It is obvious that when using DEM data only, smaller objects whose difference in elevation relative to their environment is too slight are hard to detect. Within rural areas meaningful boundaries are mostly given by the relief or by the borders of settlement areas. Especially within housing areas or rural settlement areas, which hold a typical heterogeneity of elevation,[2] this effect can be observed. In contrast, a segmentation based on DEM only is not suitable to detect roads – especially in rural areas. In such cases it is rather useful to take the sensor's spectral information into account. On the other hand, the DEM is well suited to detect exposed objects of interest such as buildings and houses even on large scale and within dense urban areas. Using the spectral channels when applying the top-down-approach, a behaviour of relevant image objects similar to a DEM based segmentation can be observed. However, in contrast to the DEM-based segmentation, the objects' shape now is determined by spectral properties, which leads to different objects in terms of shape and semantic meaning. Now problems occur within areas of low contrast in colour. Especially on higher levels the borders of buildings are not extracted correctly. The errors are mainly caused by shadows of larger buildings as well as by "blurring effects"

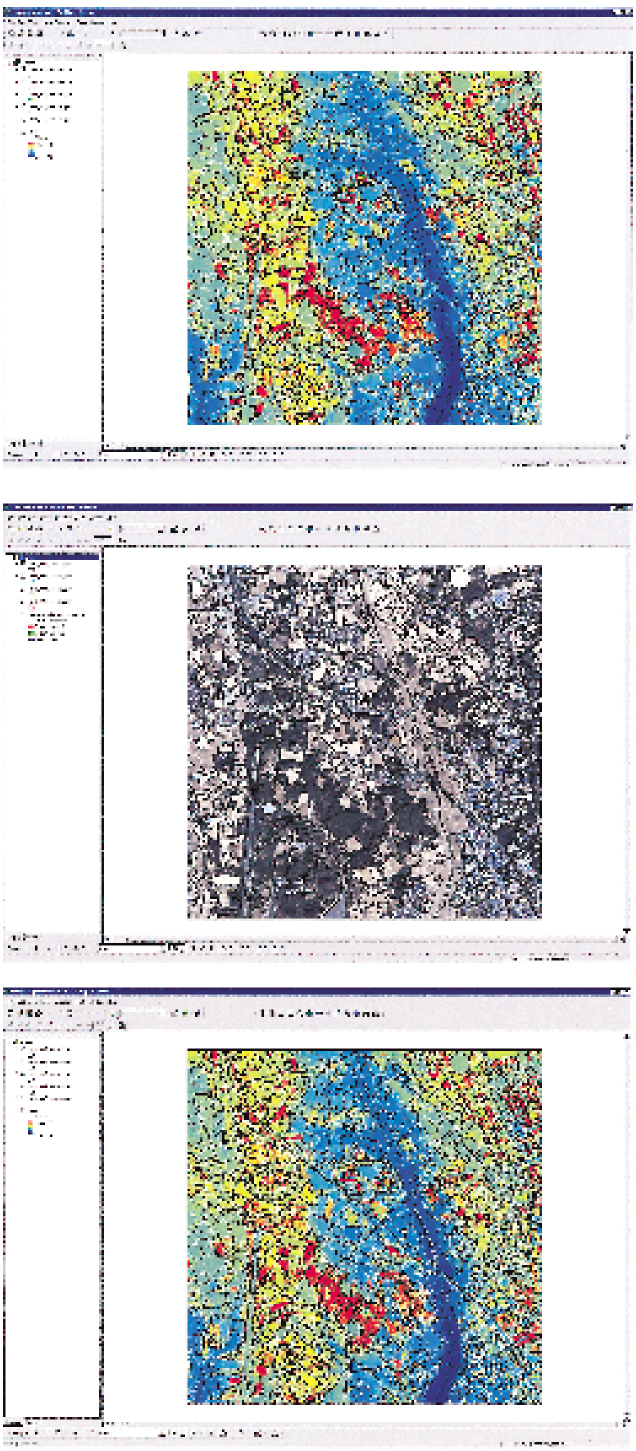

Fig. 2: Segmentation results for the highest level, based upon DEM (top), Image channels (middle) and both combined (bottom)

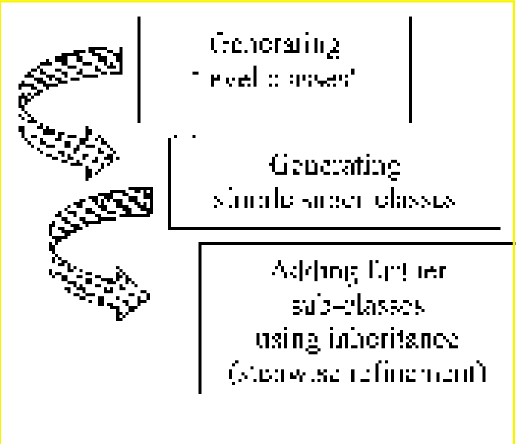

Fig. 3: General workflow for generating a class hierarchy

between impervious areas and bordering buildings. While in the first cases even the human eye is unable to detect the real shape, in the latter case a human operator would draw the correct boundary line just due to his experience. Nevertheless, relevant object borders, e.g. of roads and fields are usually generated correctly. Regarding the segmentation one has to differentiate between size on the one hand and elevation and colour properties on the other. Large and significantly high objects are best outlined focusing only their elevation properties (buildings and building complexes), while large objects with significantly high spectral contrast are better outlined taking their colour properties into account (roads, large impervious areas and fields). Medium-sized and small objects are best outlined by simultaneously regarding their spectral and elevation properties with similar weights. Thus, it is reasonable to perform the segmentation of the lower levels either by just using the spectral channels or by mixing the 'elevation channel' with the spectral channels. eCognition therefore allows to weight the channels' influence on the segmentation by appropriate adjustment. Additionally, during the segmentation process, the boundaries of newly generated image objects follow the boundaries of still existing sub- or super-objects. Hence, the order of the segmentation crucially affects the object generation.

Taking this sensitivity of the results into account, a combined top-down-/bottom-up-approach leads to most reasonable image objects for the following reasons:

- Many large higher-level image objects represent essential objects of interest (buildings and building complexes). Thus, their boundaries must not be destroyed or falsified by a finer segmentation. They act as supreme boundaries for the subsequent segmentations.
- Small objects hold useful detail information for the subsequent classification, regardless whether they will be classified or not (e.g. for texture description). If they are classified, they can be merged to more meaningful medium-sized objects (classification-based segmentation).
- Medium-sized objects ideally represent a reasonable aggregation of their sub-objects. Simultaneously they are embedded into a larger context. Hence, they reflect the image content more detailed than their coarser high-level objects and hold more meaningful semantic and shape information than their sub-objects.

Fig. 4: Example for an initial class hierarchy with simple super-classes

As in eCognition the image segmentation works on the entire image, segmentation in the present example was performed as following:

a) Initial fine segmentation giving the DEM a strong weight (parameters see *table 1*). Note that a zero-weighting of the spectral channels leads to meaningless borders in flat regions *(figure 2)*.
b) Coarse DEM-based segmentation with zero weights for spectral channels.
c) Intermediary segmentations, with and without weighting the 'elevation channel' while 'compactness' receives a stronger weight.

Classification

Once image segmentation has been performed, the image segments are classified by generating a class-hierarchy. Within this class-hierarchy each class is described either by one or more fuzzy-membership functions, a nearest neighbour classifier or by a combination of both[3]. Using the inheritance mechanism, a stepwise refinement of the class-hierarchy can be achieved. Generating classes for each level helps to separate the content of the different image levels[4]. This is because all objects contained in each level will inherit the property of the respective level, leading to a more transparent and clearly structured class-hierarchy. Following the refinement concept, in the beginning simple classes are generated which are easy to describe (in the most cases by a spectral nearest neighbour classification). These classes act ideally as semantic super-classes (e.g. vegetation as a super-class of meadows and forests). The sub-classes then have two types of properties: inherited and their very own, specific properties. The typical workflow of generating a class-hierarchy is demonstrated in *figure 3*.

---

[2] This information is very helpful for subsequent classification

[3] By means of fuzzy-logic operators

[4] The only property of each "level class" is that it belongs to the respective level

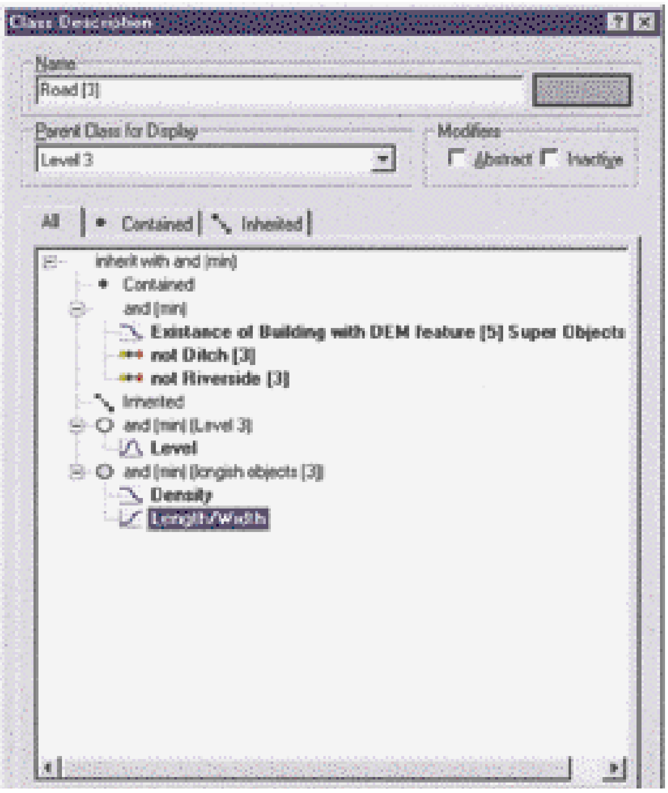

Fig. 5: Class description of roads

Classification based upon spectral properties only

Since the generated image segments hold more spectral information compared to pixels' digital numbers, eCognition offers a huge variety of derivative spectral features[5]:

- Brightness, spectral ratios.
- Textural features: standard deviations of channel values, spectral mean values of sub-objects, average spectral differences of sub-objects.
- Contrast features: spectral differences to neighbouring objects, super-objects, or the entire scene, relative border to brighter neighbours.
- Context related features: Mean spectral differences to a given class.

For most simple classes spectral nearest neighbour classification is sufficient, particularly for object classes of the highest level. In the present example the initial class hierarchy is shown in *figure 4.* A fuzzy membership description of the standard deviations in the near-infrared channel (IKONOS channel 4) for the classes settlement areas and dark field improved their classification. Nevertheless some fields, impervious areas and dark buildings were misclassified which indicates a better separability of these classes using DEM-based features or other non-spectral features.

Classification based upon spectral and DEM-properties

When taking DEM-information into account, it is obvious that it is not the absolute but the relative height within classes which characterises them. In eCognition this property can be modelled by describing the difference in elevation to neighbouring objects, which is comparable to the contrast features in the chapter above. Furthermore the standard deviation within the 'elevation channel' discriminates flat areas and "variably surfaced" areas (e.g. settlement areas, forests vs. fields and meadows). In the example, adding such features improves the classification as follows:

- Buildings and shadows are better detected by their relative height differences between neighbouring objects.
- The classification of settlement areas and forests was improved by describing their high variation in elevation.
- Fields, meadows and larger impervious areas were better detected describing their low variation in elevation.

Nevertheless some objects still remained mis-classified, which can be explained as follows:

- Road segments on bridges are classified either as dark buildings or as forest, because bridges are objects with high elevation difference to their environment.
- Small and heterogeneous impervious areas are classified as settlement areas. Both classes are very similar in colour and elevation and thus can only be distinguished by considering their context.
- Some fields are classified as impervious areas and vice versa. This is mostly explainable by the time of acquisition, since on the northern hemisphere in March agricultural vegetation is mainly inactive. Hence, many fields look very similar to larger impervious areas.

With respect to the classification task roads are hardly detectable considering their spectral and elevation properties only. Since they are typically long and narrow, shape criteria are better suited to describe them.

Classification enhancements usingform and context features

As roads are typically elongated features, most of them can be detected describing just their shape *(figure 5)*. Additionally, roads may vary in their spectral properties. Thus it is useful to classify roads by describing their shape by form criteria and subsequently their different spectral properties. Therefore, depending on the type, roads may inherit their spectral properties from an appropriate super-class (e.g. from dark impervious areas or from dark fields) and then be identified by their form. Alternatively, a road super-class (e.g. elongated objects) holds the common shape properties and the different road-types are described by their spectral properties. Either way, several road-classes with different spectral but similar shape properties are obtained, which can be combined into one semantic road-group. In the present example a ditch is running from north to south, which looks very similar to roads. To differentiate it from roads, its absolute and relative elevation, its contrast in the NIR channel and its direction has been used. To avoid mis-classification of the border-

[5] The next release will enable the user to combine existing features arithmetically. The typical ratios like the NDVI can be generated

ing river banks as roads, their common border with the ditch has been used. With respect to the top-down approach for classes which hold smaller objects within a larger context, the existence or non-existence of certain classes in levels above was used successfully. For instance, the class building does not occur in forest-areas and roads are not crossing buildings *(figure 5)*.

Conclusion and Outlook

Regarding the results gained from the created class hierarchy, most of the objects of interest could be identified *(figure 6)*. With respect to some contextual features and the similarity of some classes (e.g. that of impervious areas and fields), a final manual revision of the classification could not be avoided, which was performed quickly in a 'click-and-classify' manner. As the classification strongly depends on the quality of the initial segmentation, additional DEM data, especially in urban areas, can improve the analysis of high resolution image data. Specifically in cases when shadows cover objects of interest their shape can be better generated and described due to their height. Thus, for a successful segmentation and classification, geometrical shifts between the data as well as noise errors in the DEM must be avoided as far as possible. When using a DEM during the segmentation, its weight should be adjusted adequately, depending on the objects which are to be extracted. Regarding the complex context and the different extents of urban, suburban and rural objects the multi-resolution segmentation technique of eCognition combined with its object-oriented classification approach offers an appropriate tool to handle such complex information. According to an analog image interpretation, applying a classification strategy of stepwise elimination and refinement is convenient. Describing the embedding of small objects into a larger context by regarding the classification of coarser super-objects reflects a coarse-to-fine (top-down) strategy. Nevertheless buildings and roads as so called manmade features usually have a typical shape which is very often disturbed or deformed by other objects or phenomena. In cases of large and elevated objects such as large buildings, DEM helps to improve the shape generation and subsequent classification. However, in the present version of eCognition the objects' shapes can only be generated as they are 'seen' by the sensor(s) . With edge-processing algorithms (see Baatz, M. in this issue) advanced edge processing on image segments will be possible which should lead to easier handling objects' shapes. With regard to the complexity of rule-bases this should also lead to more transparent and thus more portable rule-bases.

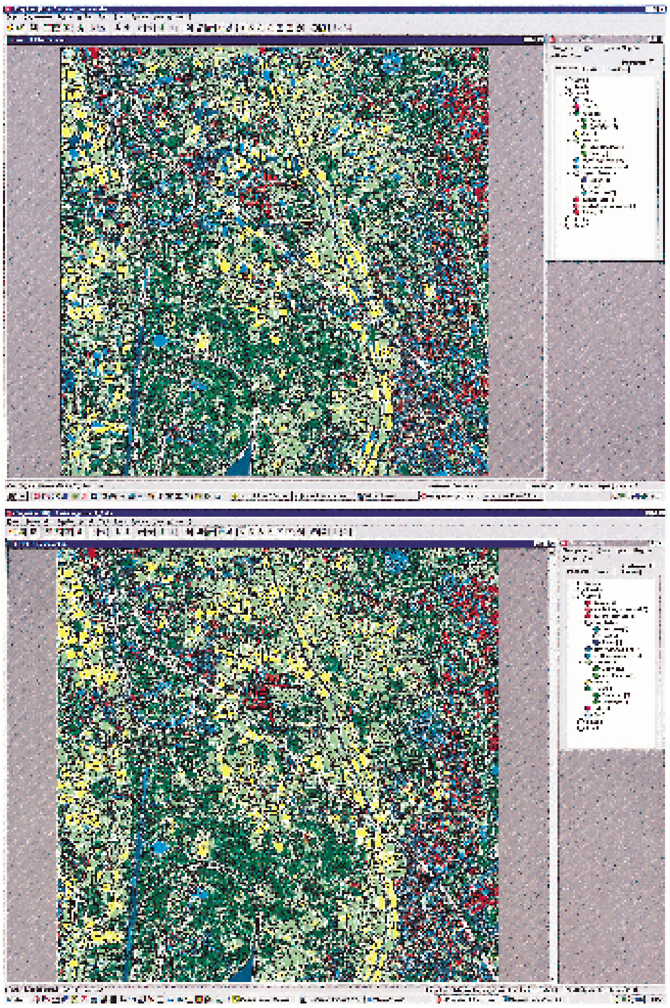

Fig. 6: Final automatic classification result (top) and enhanced revised classification (bottom)

Acknowledgement

The author likes to thank Mr. YunQuin Li and Mr. Akira Watanabe, both from Space Imaging Japan, for giving us the permission to publish about the theme, to provide the image and DEM data and their patience while I processed the data.

References

BAATZ, M., SCHÄPE, A., 1999a: GIS-Aktualisierung und Image Fusion basierend auf einem neuen Verfahren zur objektorientierten und multiskalaren Bildanalyse. In: Strobl, J., Blaschke, T. (Hrsg): Angewandte Geographische Informationsverarbeitung XI. Beiträge zum AGIT-Symposium Salzburg 1999. Karlsruhe. Herbert Wichmann Verlag: S. 12–24.

BAATZ, M., SCHÄPE, A., 1999b: Object-Oriented and Multi-Scale Image Analysis in Semantic Networks. In: Proc. of the 2nd International Symposium on Operationalization of Remote Sensing August 16th – 20th 1999. Enschede. ITC.

BÜCKNER, J. ET AL., 1999: Knowledge Based Interpretation of Objects in Topographic Maps and Moorlands in Aerial Images. In: Förstner, W. et al. (Edtrs.): Semantic Modeling for the Acquisition of Topographic Information from Images and Maps. SMATI 1999, p. 25 – 40.

DEFiNiENS AG, 2000: eCognition User Guide. http://www.definiens.com/register/index.html.

DEKOK, R., ET AL., 1999: Object based image analysis of high resolution data in the alpine forest area; Joint WSf ISPRS WG I/1, I/3 and IV/4: SENSORS AND MAPPINHG FROM SPACE 1999; Hannover, September 27–30, 1999.

HINZ, S. ET AL., 1999: Road Extraction in Rural and Urban Areas. In: Förstner, W. et al. (Edtrs.): Semantic Modeling for the Acquisition of Topographic Information from Images and Maps. SMATI 1999, p.133 – 154.

HOFMANN, P., REINHARDT, W., 2000: The extraction of GIS features from high resolution imagery using advanced methods based on additional contextual information – first experiences. In: IAPRS, Vol. XXXIII, Amsterdam, 2000.

KNOKE, R., 1999. Extraktion objektstrukturierter Geoinformationen aus Fernerkundungsdaten, unpublished diploma thesis, Univ. of federal armed forces Munich.

LIEDTKE, C.-E. ET AL., 1997. AIDA: A system for the knowledge based interpretation of remote sensing data. http://www.tnt.uni-hannover.de (July 1997).

WILLHAUCK, G., 2000. Comparison of object oriented classification techniques and standard image analysis for the use of change detection between SPOT multispectral satellite images and aerial photos. IAPRS, Vol. XXXIII