THE EXTRACTION OF GIS FEATURES FROM HIGH RESOLUTION IMAGERY USING ADVANCED METHODS BASED ON ADDITIONAL CONTEXTUAL INFORMATION – FIRST EXPIERIENCES

Peter HOFMANN[*], Wolfgang REINHARDT [**]
[*]Delphi2 Creative Technologies, Germany (since January 2000, former AGIS, Universität der Bundeswehr München)
Peter.Hofmann@delphi2.de
[**] AGIS, Universität der Bundeswehr München, Germany
Wolfgang.Reinhardt@unibw-muenchen.de

Working Group IV/6

KEY WORDS: Classification, Fuzzy, GIS, High-resolution data/images, Object-oriented.

ABSTRACT

Extracting GIS features from high resolution imagery for map production or to update or generate GIS databases will be a common task in remote sensing in the coming years. These features are often described and defined in feature catalogues such as the German ATKIS or the military Vmap catalogues. Using sensors of lower resolution such as LANDSAT TM or SPOT only features relevant for small scale could be extracted. Methodically therefore it is sufficient to use conventional statistically based classification methods for feature extraction. The usage of high resolution image data such as scanned orthophotos or data taken from high resolution air- or spaceborne scanners, makes it possible in principle to extract features relevant in large scales. But this also requires advanced methods for feature extraction because in these cases, conventional methods of image processing will fail very often. One approach to improve this is to use methods which are able to include additional information on the features like its shape or context. In general that means, methods for image analysis are needed which are closer to human visual perception.
A problem with the use of additional feature information in image analysis is that this information is hardly describable in a sharp manner. One solution for this problem could be to describe it in a fuzzy way. The methods used in this investigation also take account of the features different appearance in different image resolutions. Hence appropriate segmentation algorithms as well as classification methods which take account of the features spectral and non-spectral behaviour mentioned above are necessary.

1 INTRODUCTION

The aim of this paper is to give our first experiences of how and respectively which kind of additional information can improve the process of feature extraction when using high resolution multispectral airborne scanner data (here: DPA-data from Bückeburg, Germany). The features we try to extract are taken from the Vmap level 1 feature catalogue and from the ATKIS DLM 25 catalogue. For image analysis we use the new software *eCognition* by *Delphi2* whose methodology we will roughly outline. A final comparison between results from conventional image processing methods like spectral pixel classification and the used methods offered by *eCognition* will be given.

2 DIFFERENCE BETWEEN IMAGE- AND GIS-FEATURES

While in digital images a certain part of the earth's surface at a certain moment is pictured and stored in an array of pixels in which only gray values are stored, a map or GIS data shows the same content in an abstract and generalized form, such as e.g. in a map. Hence the appearance of the features to be extracted from the image depends strongly on the image's parameters (such as date, location and sensor) and the feature's inherent appearing in the image (such as color, texture and shape). Additionally the detection of GIS or map features from an image can be enhanced by considering the context in which a feature could be embedded or not. Although the human eye and the human ability of perception is capable in general to detect (GIS-) features even in complex situations or distorted images as well as within different contexts, an automated computer based process which achieves comparable results as a human visual image interpretation is not to be expected within the coming years. Nevertheless remote sensing data is a useful data source to generate or update GIS data bases mainly because of its high actuality. Especially for large scale GIS-data only image data of high resolution such as scanned aerial photographs, airborne scanners or satellite based scanner data of the newer generation depicts an adequate data source. On the other hand making high resolution remote sensing data manageable, depends strongly on the development of adequate methods to detect and extract features which are commonly kept in GIS data bases. Recent research in the fields of semantic modeling and computer vision gives promising but proprietary solutions (Janssen, L.L.F. & Middelkoop, H. (1992); Baumgartner, A. et. al. (1999); Newby, P.R.T. (1996); Liedtke, C.-E. et al. (1997); Tönjes, R. & Growe, S. (1998)).

Since DELPHI2 developed a new software named eCognition which is scheduled to be available on market in July 2000, new approaches of image segmentation and image analysis are united within one software package. For a better understanding of the strategies we tested by using eCognition, we give a short description of the methodical principles of that software. A more detailed description of the software and its segmentation methods implemented in are given in Baatz, M. & Schäpe, A. (2000) and Schmidt, R. (2000).

3 IMAGE DATA AND FEATURES TO BE EXTRACTED

In the described experiences image data derived from the DPA sensor from Bückeburg has been used. Table 1 gives a short overview of the image's characteristics.

Date	Altitude	Pixel size on ground	Spectral sensitivity
11th July 1997 12:30	10224ft (3116m above sealevel)	0.75m	Channel 1 (blue) 0,44 – 0,53 µm
			Channel 2 (green) 0,52 – 0,60 µm
			Channel 3 (red) 0,61 – 0,69 µm
			Channel 4 (near IR) 0,77 – 0,89 µm

Table 1 DPA-Data characteristics

For the further tests four areas have been token from the whole scene which cover different types of landuse: Two rural areas with typical Mid-European rural settlement structures and landuse forms, one test area covering urban and suburban areas of Bückeburg and one test area covering Bückeburg airfield. The experiences of our tests as well as the

Figure 1. Rural test area token from the DPA scene channels 1 (blue), 2 (green) and 3 (red)

methods and strategies we used are demonstrated using two rural areas such as shown in figure 1.
From the test areas the following GIS features according to the feature catalogues of VMap resp. ATKIS had to be extracted as far as they could be visually discovered in the test areas:

arable farm land	grass land	forest	single trees
waters	tarred roads	untarred roads	parking site
untarred pathways	field paths	roadside green	sports field
buildings	gardens	grave yard	bridge

Table 2. Features to be extracted from DPA test areas

4 METHODICAL PRINCIPLES OF eCognition

In contrast to conventional image processing software which is mainly based on pixel statistics, eCognition deals with image objects. Image objects within eCognition are derived by using a new patented segmentation technique developed by DELPHI2 (Baatz, M. & Schäpe, A. (2000)). The segmentation process thereby takes into account the homogeneity in form and color of the objects to be generated as well as their size. The generated contiguous image segments then should ideally represent real world objects on different levels of scale. Additionally the created segments are embedded into a hierarchical network wherein each segment (resp. object) 'knows' its neighbors (fig. 2).

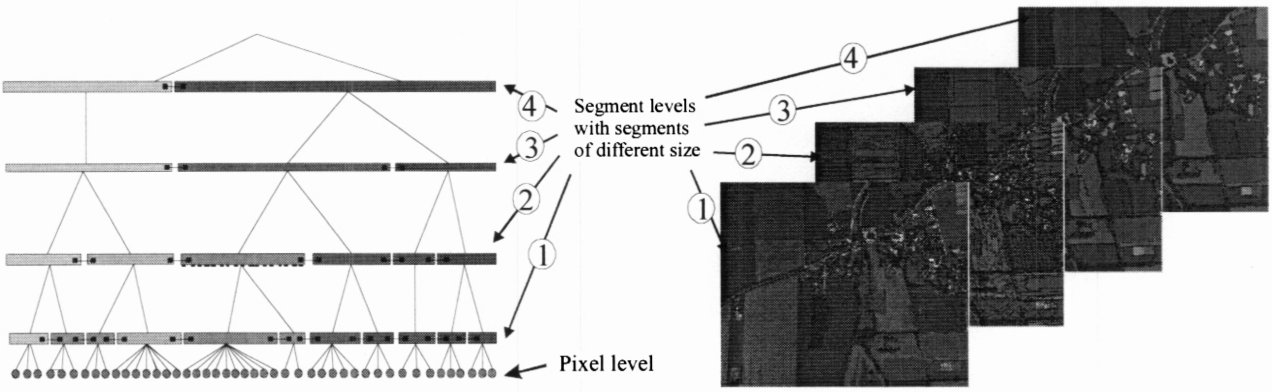

Figure 2. Hierarchical net of image objects derived from image segmentation

The image objects then can be described and classified by a huge variety of features which include color-, texture-, form- and context properties in several forms. This is either be done using a nearest neighbor classifier or fuzzy membership functions (fig. 3).

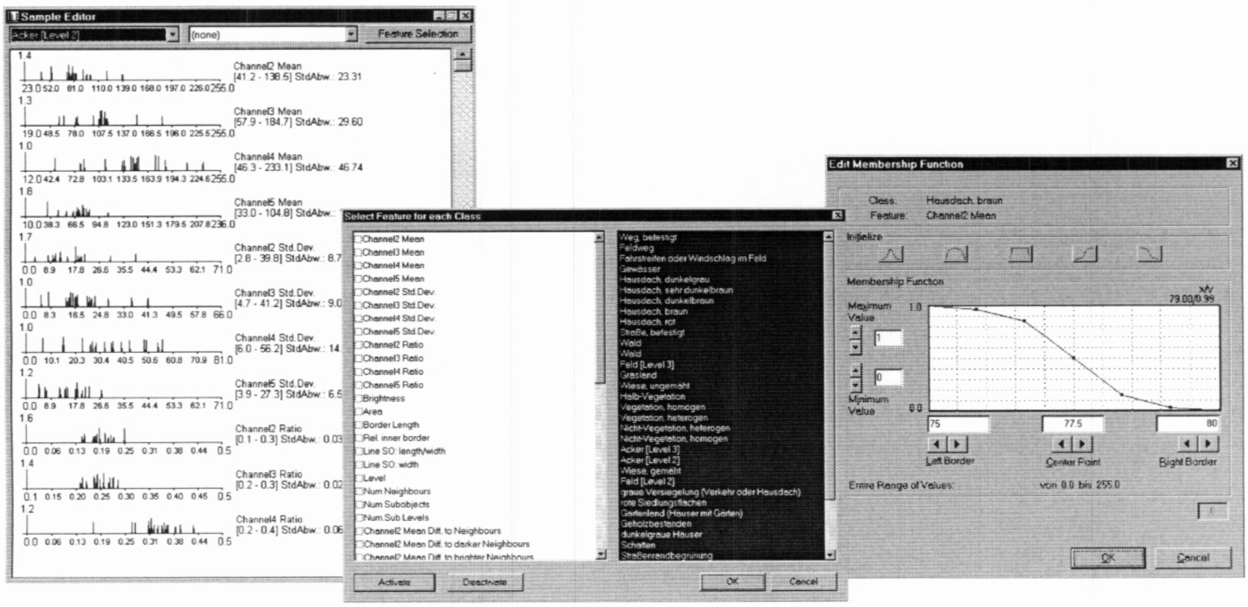

Figure 3. Class description using nearest neighbor classifier or fuzzy membership functions

The description of the classes, their entities and relationships is done within a class hierarchy which simultaneously acts as a knowledge base describing the image's content. Within the class hierarchy it is possible to create sub- and super classes, which allows it on the one hand to inherit feature properties from super- to sub-classes and on the other hand to group classes semantically (fig. 4).

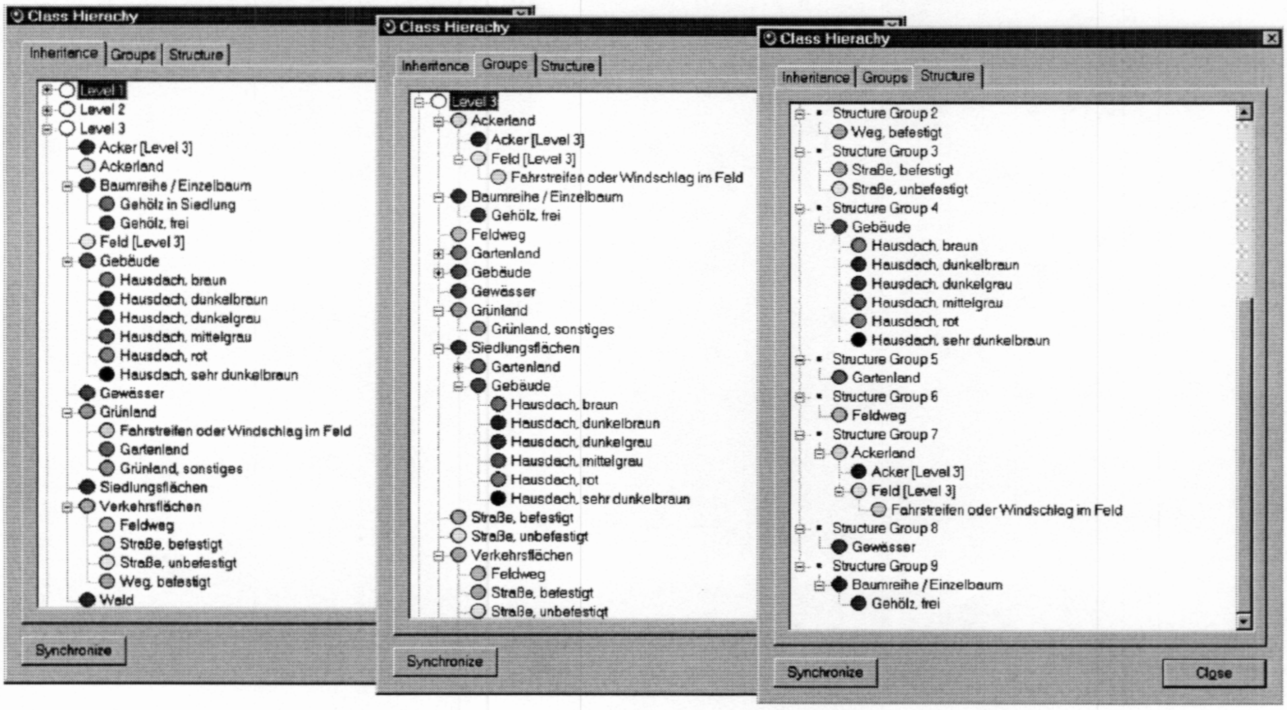

Figure 4. Class hierarchy with property inheritance and semantic grouping.

Creating structure groups allows it to eliminate borders of spatially neighbored segments of the same class or to create a new level which only contains contiguous segments of one semantic group. Creating structures can also be understood as a knowledge based segmentation.

5 CLASSIFICATION STRATEGIES

According to the multiscale hierarchical network created during the segmentation process, features are best extracted, if the image segments represent the feature's shape, size and color as good as possible. Thus some features are better represented by segments of a higher segmentation level, while others are better detected on a lower level. Hence we decided to investigate two basic classification and processing strategies: On the one hand a **bottom up approach** which means that beginning with small segments, larger segments in the levels above should be generated and classified. On the other hand a **top down approach** where at first large segments are generated and roughly classified, smaller segments on lower levels have been generated and classified whereas the classification of the smaller segments than refers to the classified segments in the levels above.

Additionally we decided to use only one level within the segment tree as the level containing the final classification results. Usually this should be a level where the smallest detectable features to be extracted are best represented by the segments derived from the segmentation process. Further it has to be considered that the description of texture and form in eCognition is based on size and shape of sub-segments within the segment tree. This means that for making texture and form usable as describing properties, at least one segment level below the level to be evaluated must exist. Thus we generated an appropriate segment level, containing only small and mostly meaningless objects. As classification methods we used for both approaches a combination of spectral nearest neighbor classification and fuzzy descriptions of the generated texture and form features as well as for typical spectral properties for the classes the segments should belong to.

5.1 Spectral classification of segments

Due to classifying homogeneous image segments, the segments inherent spectral information is represented by parameters such as: *mean gray values* in one or more channels, *standard deviation of the gray* values in one or more channels or *proportions of the color mixture* of the segment. Thereby starting with a spectral nearest neighbor classification of the segments based upon the mean gray values of the DPA channels 1, 2, 3 and 4 and the standard deviations in the same channels gave us a first rough classification result. Misclassifications occurred according to the spectral similarities of the features to be extracted (fig. 5). While some feature classes (such as *sports field*) could be found directly, other classes (such as *arable farm land* or *roof*) could be detected hardly (Knoke, R. (1999)). After this first classification step, using abstract class names has been experienced as very useful, if there obviously misclassified segments within a class appeared. Class names such as *Similar to arable farm land* or *Similar to roof* were chosen for small segments in the bottom up approach. In the top down approach names like *Semi-Vegetation*, *Non-Vegetation* and *Vegetation* have been chosen. This technique has led to avoid misclassifications resp. to gain useful classes which could act later as abstract superior classes. Due to the almost insufficient results of the nearest neighbor classification, further feature inherent properties have been described by fuzzy membership functions (fig. 6 and 7). Properties describing the typical appearance of the features in their size, color or texture thereby have been chosen individually.

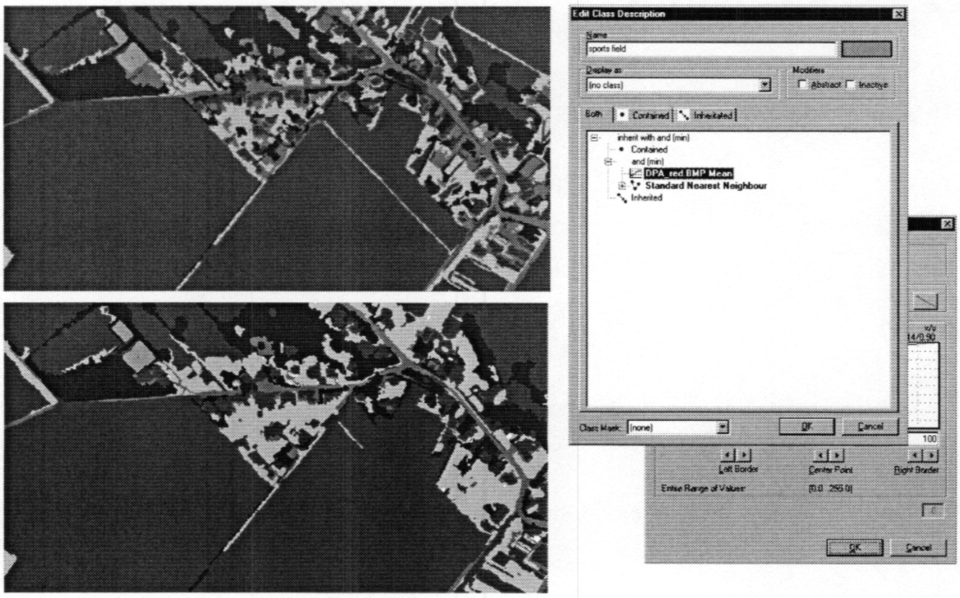

Figure 5. Results of spectral classification in the Bottom Up Approach. Top: Results from nearest neighbor classification. Bottom: Enhanced by additional fuzzy membership functions. Right: Membership function of *Sports Field* for mean gray

5.2 Inheritance

Additionally sub-classes of the abstract classes derived from the spectral classification have been created. This can be understood as instantiating the sub-classes from superior classes. Thereby the occurring sub-classes are inheriting the spectral entities but differ in typical feature-inherent properties (mostly form, size, texture or individual gray values). To formulate the characteristics of a class inherent property, fuzzy membership functions have been used. Usually the used properties thereby should distinguish the features resp. sub-classes as good as possible (fig. 6). E. g. the difference of forests and trees lies in their size resp. scale in which they appear as features. But both classes can be understood as instances of wood vegetation which inherits its spectral entities. It has to be emphasized, that only physical properties are inherited during this stage.

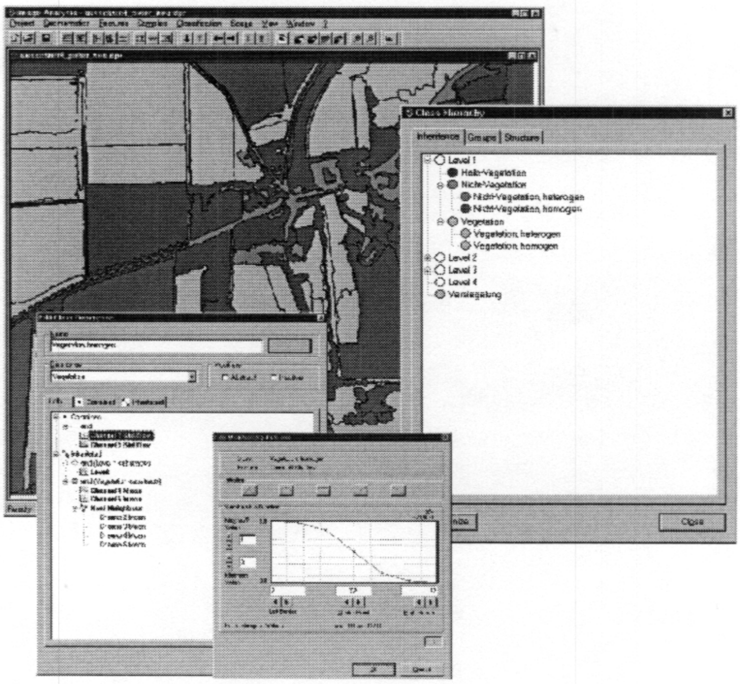

Figure 6. Inheriting subclasses *(Non-)Vegetaion heterogeneous/ homogenous* as instances of *(Non-)Vegetation* in the Top Down Approach.

5.3 Using context information

Context information in eCognition can be described and used in two principal ways: On the one hand as spatial context, which means neighborhood entities of the classified segments within the image tree in horizontal and vertical direction is described and on the other hand as semantically context by grouping the classes to reasonable semantic groups, which means that despite physical similarities to a class a segment belongs to a semantically completely different class (fig. 4). It has to be considered, that the segments' physical and semantic classification influences their spatial context. Therefore at least the physical classification of the segments should be as proper as possible. Due to the two approaches contextual information has been used in different ways: In the top down approach it was essential for spatially large features (such as *arable farm land* or *grass land*) to use spatially vertical context information. Large correctly classified segments of higher levels then determined the assignment of smaller segments within lower levels. Thus segments of the classes *arable farm land* and *grass land* in the final classification level have been classified due to their spatial relation to the larger segments in the levels above. Additionally the classes *Non-Vegetation heterogeneous* and *Non-Vegetation homogeneous* of the top level have been used to distinguish between *settlement areas* and *arable farm land,*

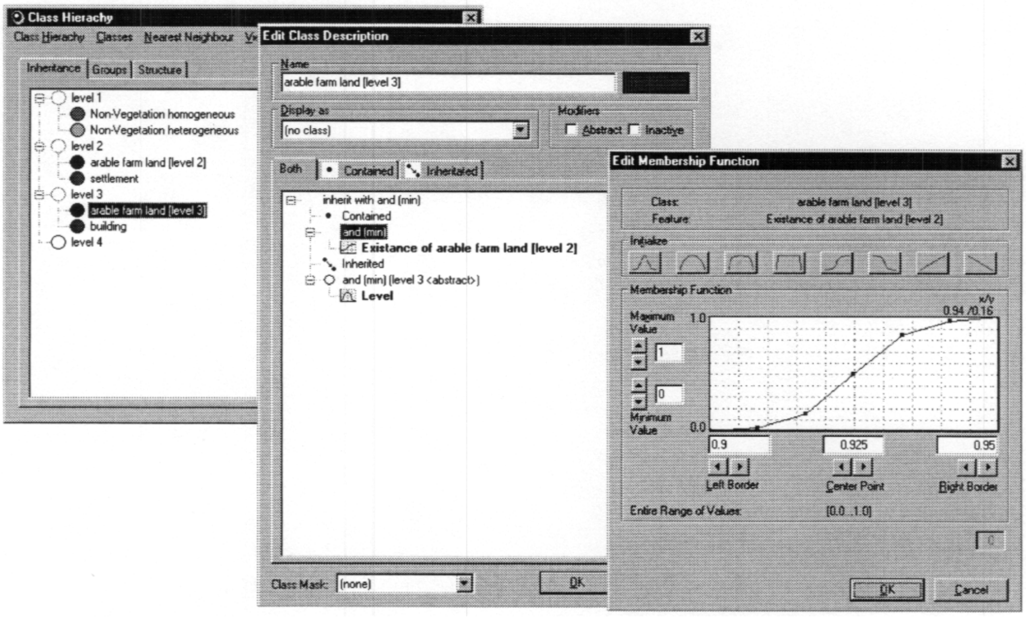

Figure 7. Example for using vertical context classification.

which in the resulting classification level led to a better determination of *buildings*. While in the pure spectral classification smaller segments within *arable farm land* areas have been misclassified as *buildings,* those segments have been classified correctly by using the vertical context information of *existence of arable farm land* resp. *existence of Non-Vegetation homogeneous* which acted as a superior class of *arable farm land* (fig. 7).
In the bottom up approach mainly horizontal context information has been used to enhance the physical classification. Thereby enclosed misclassified segments have been eliminated by creating intermediate classes like *similar to arable farm land* then creating a sub-class named *grass land similar to arable farm land* which finally inherits the spectral entities of *similar to arable farm land* but has a significant longer common border with *grass land*. Semantic groups have been created in cases when higher aggregated features could not be detected directly. Especially buildings could not be detected as a whole. Furthermore several forms of roofs differentiated by their colors have been found. But for the final result only the class *building*, which groups the different colored roofs is relevant (fig. 4).

6 FINAL RESULTS AND COMPARISON TO PIXEL BASED METHODS

Due to the features that should have been detected according to the feature catalogue, the extracted real world features found in the image have been aggregate to features which appear in the feature catalogue. Figure 4 therefore gives an idea, how this can be done in eCognition. In figure 8 a final result we derived from the segment based classifications with eCognition is shown. Compared to a pixel based classification using a supervised maximum likelihood classification and a manually visual interpretation, as a first visual impression the segment based classification is more close to the results derived from the visual interpretation. Nevertheless some misclassifications occurred in areas where even the human eye can hardly distinguish the features, such as the agricultural area in the center, where it is not sure whether it is *grass land* or *arable farm land.* Compared to the supervised maximum likelihood classification the segment based classification gives a more homogeneous impression, which is due to the prior segmentation process. In both classifications the very strong textured *grave yard* in the left center has not been detected, but while in the pixel based classification the neighbored field has been mostly misclassified as *grave yard,* this has not occurred within the segment based classification. The detection of roads was not sufficient enough in the pixel-based and the segment based classification. But using form properties in eCognition helped to describe their typical linear appearance. The relatively well detected *gardens* in the segment based classification is a result of using additional context information. While in the pixel based classification larger *grass land* areas outside the settlement area are classified as *gardens*, this could be avoided by formulating adequate fuzzy membership functions describing the relative neighborhood of *gardens* to *buildings* resp. *settlement areas.* Compared to the visual interpretation the pixel based classification as well as the segment based are lacking a sufficient geometrical application of the features to be detected. While the pixel based classification shows many distorted areas, the shape of some classified segments is unsatisfying.

7 CONCLUSIONS

Detecting GIS features directly from high resolution remote sensing data will still be a challenge in the coming years. Since the information contained in an remote sensing image can be very complex and is very different to that kept in a GIS, the content of an image can be displayed in a sketch similar to a map. As in the paper shown newer approaches such as the segment based classification can improve the image classification results increasingly. Especially the segment based approach makes it possible to formulate more complex entities and relations of the objects within an image such as form or context. A further improvement compared to conventional classification methods is, that image

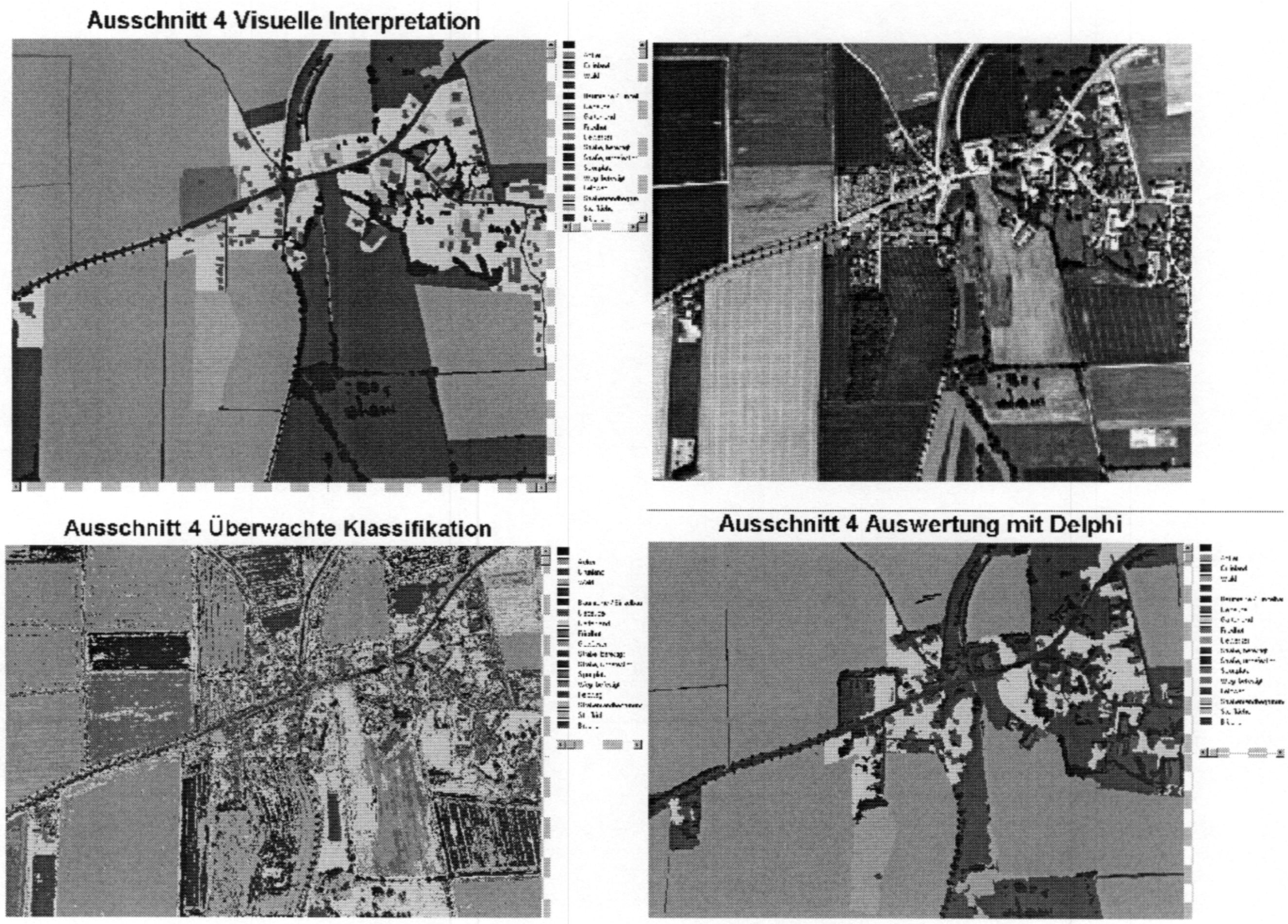

Figure 8. Results derived from a supervised maximum likelihood classification (bottom left), a visual interpretation (top left) and the segment based classification (bottom right).

data can be analyzed on different levels of scale which takes into account the different appearances of the features according to their size. Though the top down approach gives a more secure classification result due to its higher transparency, the bottom up approach has been experienced as the faster method. Finally it has to be retained, that the techniques of inheritance and semantic grouping can be used to formulate the complex situations of high resolution images in an intuitive and compact way. Thereby we experienced as well, that for gaining concrete features of a feature catalogue it is more useful as a first step to extract real world features that can be seen in the image and then to merge them to classes of the feature catalogues. Choosing abstract class names which more describe the segment class' appearance in the image therefore is an appropriate instrument. Thus the description of the image's content is described in the final results' class hierarchy which simultaneously acts as a semantic net which describes the pictured real world objects in their appearance, neighborhoods and their context. Thereby the class hierarchy can be understood as a semantic model of that part of the world, which is pictured in the image. From this model finally the features of the catalogues to be extracted then can be generated easily by assigning the real world objects to the appropriate feature classes. It has to be mentioned, that using context information can become serious, if the classes or semantic groups on which the contextual information is referenced to, are not classified adequately. Thus a prior classification of the segments based on their physical appearance should be as good as possible. Due to our experiences made and those described by Willhauck (2000), eCognition and its technology seems to be very promising. Since the principles of the segmentation algorithm is published in June 2000 further more scientific investigations are possible and should from our point of view be done.

REFERENCES

Baatz, M. & Schäpe, A., 2000. Multiresolution Segmentation: An optimization approach for high quality multi-scale image segmentation. In: Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT Symposium Salzburg 2000. Wichmann, Karlsruhe.

Baumgartner, A. et. Al., 1999. Automatische Straßenextraktion auf Grundlage von verschiedenen Auflösungsstufen, Netzbildung und Kontext. Photogrammetrie Fernerkundung Geoinformation (PFG) 1/1999, pp. 5-17.

Janssen, L.L.F. & Middelkoop, H., 1992. Knowledge-based crop classification of Landsat Thematic Mapper image, in: Int. J. Remote Sensing, Vol. 13, No. 15, pp. 2827-2837.

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).

Newby, P.R.T., 1996. Digital images in the map revision process, in: ISPRS Journal of Photogrammetry & Remote Sensing 51, p. 188-195.

Schmidt, R., 2000. Untersuchung des Bildanalysesystems eCognition. Unpublished diploma thesis Univ. Hannover.

Tönjes, R. & Growe, S., 1998. Knowledge based road extraction from multisensor imagery. http://www.tnt.uni-hannover.de/~toenjes, ~growe (July 1998).

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