Potential and problems of multi-scale segmentation methods in remote sensing

ABSTRACT

Region-based approaches are suitable for the analysis of high-resolution remotely sensed data. In this context, special emphasis has to be laid on the segmentation of multi-sensoral and multi-scale input data with respect to the geometrical and semantical robustness of the achieved results, but also on the grade of automatization and transparency of the algorithms.

ZUSAMMENFASSUNG

Potenzial und Probleme multi-skaliger Segmentierungs-methoden in der Fernerkundung.

Regionen-basierte Ansätze haben sich für die Auswertung hochauflösender Fernerkundungsdaten als brauchbar erwiesen. In diesem Zusammenhang muss besonderer Wert auf die Segmentierung der multi-sensoralen und multi-skaligen Eingabedaten gelegt werden, sowohl im Hinblick auf die geometrische und semantische Robustheit der Ergebnisse, als auch auf den Grad der Automatisierung bzw. der Transparenz der Algorithmen.

Dr.-Ing. Jochen Schiewe

Research Scientist and Lecturer with the Research Centre for Geoinformatics and Remote Sensing, University of Vechta (Germany)

Dipl.-Geogr. Lars Tufte

Research Scientist with the Research Centre for Geoinformatics and Remote Sensing, University of Vechta (Germany)

Prof. Dr.-Ing. Manfred Ehlers

Director of the Research Centre for Geoinformatics and Remote Sensing, University of Vechta (Germany)

Addresses of all authors:
PO Box 1553, D-49364 Vechta,
E-Mail: {jschiewe, ltufte, mehlers}@iuw.uni-vechta.de, Tel.: +49 (0)4441/15-558 (-422, -423), Fax: +49 (0)4441/15-464

1 Motivation

Remote sensing data are an important source for generating or updating GIS databases in a variety of applications. However, in the past several applications have been limited by the relatively coarse spatial resolution of the available data sources (e.g., for urban land cover classification). With the development and the employment of several advanced air-borne and space-borne high-resolution sensors the importance of remote sensing data has been even more increased because the existing limitations concerning geometrical accuracy and inherent information content have been alleviated.

The automatical classification of remotely sensed data is an essential action within the process of generating or updating GIS databases. Unfortunately, the high spatial resolution of the advanced sensors increases the spectral within-field variability and therefore may decrease the classification accuracy of traditional methods on per-pixel basis (like the Maximum-Likelihood method). After briefly discussing other alternative solutions, chapter 2 will describe and prove the need for a region-based approach that performs a segmentation of the imagery prior to the classification stage.

While the classification procedures are based on well-known standard algorithms, the general goal of our investigations will be to demonstrate the potential as well as the problems of segmentation methods from a theoretical and from a practical point of view. Firstly, a critical overview of existing implementations for the purpose of evaluating remotely sensed data will be given (chapter 3). Secondly, the results of our empirical studies which are aiming at an improvement of the segmentation process by introducing multi-sensoral and multi-scale data will be presented

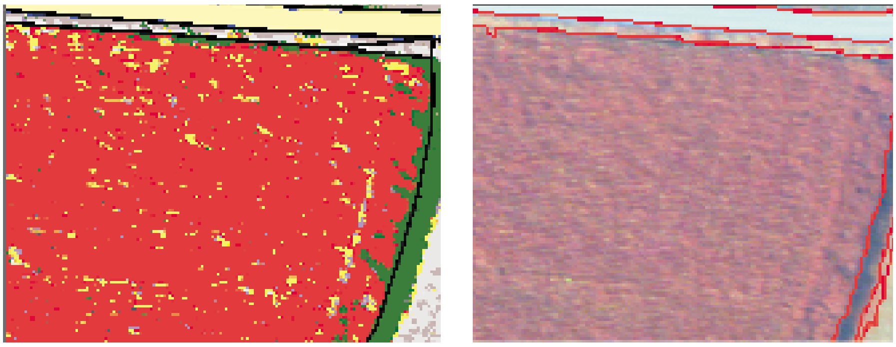


Fig. 1: More homogeneous and realistic segmentation for landuse classification purpose based on region-based method (right) compared to traditional Maximum-Likelihood method (left) (subset of 80 x 60 m²)

(chapter 4). Here, not only the geometrical and semantical robustness but also the grade of automatization and transparency of the used algorithms are discussed. Based on these experiences we will derive further research and development tasks (chapter 5).

2 Need for region-based analyses

With the development and employment of sensors with improved spatial resolutions the problem of mixed pixels is indeed reduced, but the internal variability and the noise within landuse classes are increased. As a consequence, traditional classification approaches like the Maximum-Likelihood method are producing too many or not well defined classes, because their clusters are built upon spectral homogeneities only. This general effect which is well known in literature (e.g., Cushnie 1987) is illustrated in *figure 1* (left).
As an alternative, per-field or per-parcel classification methods have been introduced. While investigations using data of medium resolution show some improvements concerning accuracy and interpretability (e.g., Aplin et.al. 1999, for a pixel size of 4 m), they have not delivered reasonable results for data of very high resolutions (e.g., Hoffmann et.al. 2000a, for a pixel size of 0.15 m).
The unsatisfying results of traditional and alternative approaches have already led to the employment of region-based approaches which consist of a segmentation of the input data prior the classification stage. *Figure 2* depicts the overall work flow which will be applied in our following investigations. While traditional multi-spectral algorithms only consider spectral similarities independently from their occurrence, the segmentation approach follows the hypothesis that neighbouring image elements belong to the same class. This hypothesis is verified or falsified on the basis of homogeneity or heterogeneity parameters. An in-depth overview is given by Haralick and Shapiro (1985) and Jähne (1997). If the segmentation is performed at multiple scale levels, optimal spatial resolutions for the different sizes, forms and arrangements of the objects under investigation can be considered. As a result one obtains connected regions (segments) – without semantical meaning so far – which in general are more homogeneous compared to results of traditional approaches *(figure 1, right)*.
The assignment of semantics to the segments has to be performed in the following classification stage through additional knowledge, for instance through an a-priori definition of an object catalogue which is adjusted to the current purpose. From our point of view, a fuzzy logic approach allowing for partial memberships is a reasonable classification technique due to the limited resolutions and accuracies as well as the potentially contradictory information coming from different data sources. Thus, an interpretation key using combinations of information – like normalized Digital Surface Model (nDSM), spectral texture, area size or neighbouring classes – has to be translated into rules and corresponding membership functions.

3 Segmentation algorithms for remote sensing data

While the classification procedures are still based on well-known standard algorithms, our following considerations and examinations will concentrate on the segmentation part. Here, numerous methods have already been developed for various applications, including medicines (e.g., Handels 2000), telecommunication engineering (e.g., Wesfreid and Wickerhauser 1999) or the analysis of dynamic scenes in neuro-informatics (e.g., Handmann et.al. 1998). Consequently, the available software products are originating from these or similar disciplines, such as NIH Image (U.S. National Institute for Health), ARIES (University of Washington, Dept. of Electrical Engineering), or DIAS (University of Jena, Dept. of Digital Image Processing).
Several reasons lead to the fact that existing methods and implementations from these disciplines can not be transferred to remote sensing, in particular:

- Remote sensing sensors are producing multi-spectral, sometimes also multi-scale input data, so that in contrast to the most often used panchromatic and monoscopic image data in the disciplines mentioned above not only the complexity but also the redundancy increases.
- Manifold additional data (e.g., GIS or elevation data) are available.
- In contrast to other applications various objects of heterogeneous properties with respect to size, form, spectral behaviour, etc. have to be considered.
- General multi-scale evaluation tools have not been developed. Tools exist only for some limited areas, for instance for the extraction of roads from aerial imagery (Ebner et.al. 1998) or for the detection of urban structures from space scanner data (Faber and Förstner 1999).
- In contrast to other applications a model-based interpretation is much more difficult due to the heterogeneity of the inherent object classes;
- In general sub-optimal solutions have not to be considered for remote sensing applications because there is no need for real-time or dynamic evaluations.

Hence, segmentation algorithms have been introduced relatively late for the analysis of remotely sensed data (e.g., Ryherd and Woodcock 1996). As a consequence, the first commercial software packages were introduced not before the year 2000.

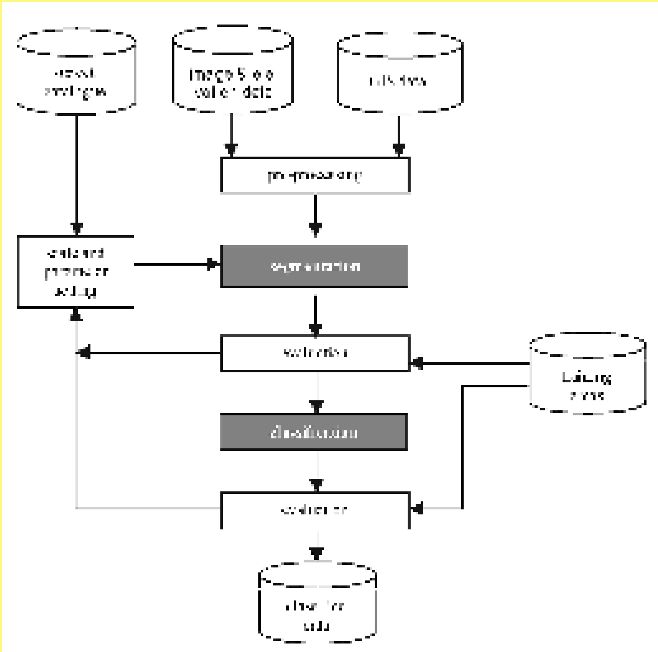


Fig. 2: General workflow of region-based multi-sensoral analysis

While special solutions are existing (like the Stand Delineation Tool of the Finnish company Arboreal for forest inventory purposes; Arboreal 2001), the software system eCognition (Definiens 2001) which will be used and discussed in the following aims at a more general use.

The underlying algorithm of eCognition joins such neighbouring regions that show a degree of fitting – computed with respect to their spectral variance and/or their shape properties – which is smaller than a pre-defined threshold (the so-called scale parameter). Eventually the choice of this parameter determines the number and size of resulting segments. Applying various scale parameters as well as different weights to the multi-sensoral data a multi-scale, hierarchical scene representation can be obtained. Finally, the classification of the extracted segments is performed using a fuzzy logic approach. Baatz and Schäpe (2000) describe the algorithms in detail.

4 Investigations concerning segmentation

4.1 Objectives

The general objective of our investigations is an improvement and operationalization of the segmentation part of the region-based analysis applied to remotely sensed data. While the approach can also be applied to other applications like for the normalization of Digital Surface Models (Schiewe 2001) we will concentrate on the land use/land cover classification. By using two test sites with different landscape characteristics (urban and rural, resp.) the emphasis will be laid on the robustness of achieved results: We look for a proper selection of input data sources and homogeneity criteria to meet accuracy and reliability measures in terms of

- geometry, i.e. the accurate positions of segment outlines, and
- semantics, i.e. the membership of one and only one class (which will be explicitly determined later on) to a segment.

Also the grade of automatization is a key issue because the acceptance of such analysis methods strongly correlates to the ability of avoiding manual tasks without loosing quality. Finally it has to be examined whether the transparency and control of the segmentation software is sufficient for an operational and flexible use.

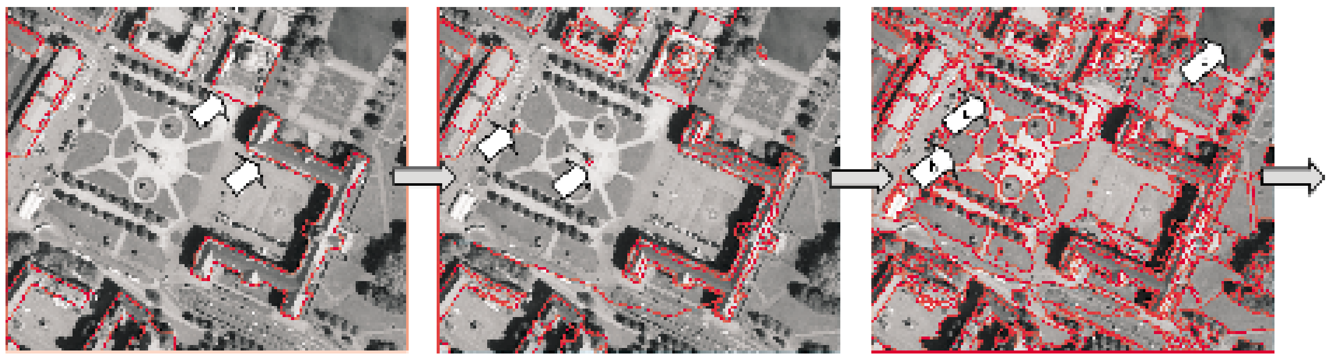
4.2 Segmentation of urban test site

The data set „Stuttgart" covers a part of the old city showing rather low terrain undulations as well as various buildings of different sizes. It consists of elevation, image and cadastral data which have not been generated simultaneously. The Digital Surface Model (DSM) has been captured with the TopoSys-Sensor (TopoSys 2001). The original measurements with an estimated vertical accuracy of about 15 cm have been transformed into a regular 1-m-grid after eliminating blunders and data gaps in a pre-processing step. Based on this a normalized DSM (nDSM) has been derived. Because only last pulse beam data have been recorded, heights of trees are not modelled. Additionally, a digital orthoimage (pan-chromatic, horizontal grid width of 25 cm) and vector data containing cadastral information (from the digital German Real Estate Map, ALK) are available.

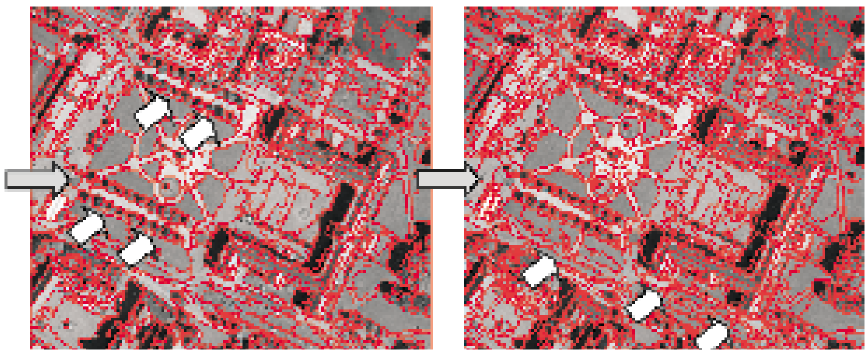
In order to follow the model-based idea as outlined in *figure 2*, the first processing step is the definition of an object catalogue containing all visible objects within the image subset – in this case: buildings, statues, roads, paths, trees, lawn, lakes. Aiming at an unique *semantical* representation within segments, it is obvious that these topographical objects are characterized best at different scales (e.g., lakes at coarser scales than trees) and through different information (e.g., buildings are separated best from other objects using their heights). As already pointed out, the hierarchical approach of eCognition is able to generate segments at different scales from coarse to fine with varying information input to the corresponding level. Although the segment outlines of coarser levels are inherited to the next scale stage, it is meaningful to keep these in mind in order to consider the most compact and less splitted representation.

Figure 3 demonstrates the result of this process emphasizing the object classes that are characterized best at the corresponding scale levels. For the first two coarse levels we use elevation information in order to extract buildings and statues. In the following only spectral information is considered in order to distinguish between other classes in the remaining regions. For our application neither a combined evaluation of elevation and image data, nor a change of weights between elevation information and shape parameters yield any significant changes in the representation of the segments. In summary, a satisfying quality with respect to the desired semantical robustness (i.e, the membership of one and only one class to one segment) can be stated. Problems arise – like in other (automatical) methods – with shadow regions and the distinction between roads and their surrounding areas.

Concerning the geometrical accuracy of segment outlines we have also obtained satisfying results which are



level 5: buildings *level 4: statues* *level 3: paths (a), lawn (b), lake (c)*



level 2: single trees *level 1: roads*

Fig. 3: Results of hierarchical and multi-scale segmentation for test site „Stuttgart" (subset: 370 x 330 m^2) with indications of object types which are characterized best on each level

mainly limited by the data quality itself. *Figure 4* – showing the overlay of segmented buildings onto the cadastral data – reflects typical values of a mean horizontal deviation of about 1 m (which corresponds to the pixel size) with maximum values of about 2.5 m. Again, it has to be pointed out that in this case the limited (horizontal!) accuracies of the laserscanning and the reference cadastral data are mainly responsible for these shifts.

A critical point within the eCognition software is the understanding of „multi-scale" processing. In fact, the software has no possibility to introduce data sets of different spatial resolutions: All data sources have to represent exactly the same subset with the same georeference and the same pixel size. Hence, the data sets have to be pre-processed, either by reducing them to a lower resolution – which means a loss of inherent information -, or by increasing to a finer resolution – which introduces undesired resampling effects. In both cases different segment outlines are produced than in the original resolution data.

In contrast, „multi-scale" processing in terms of the eCognition software means that the number and size of extracted segments can be qualitatively controlled through the so-called scale parameter (the larger the value the larger the segments). Our investigations revealed that there is an approximately linear relationship between the scale parameter and the number of segments if only one data source is considered. On the other hand, a prediction of absolute numbers for various sources is impossible. Also a comparison of a-priori spatially aggregated data and different scale parameters applied to one data set shows similar but on no account identical outlines (Schiewe & Tufte 2000). In conclusion, the segmentation is still an iterative and manual approach.

4.3 Segmentation of rural test site

The following data are based on a flight mission along a reach of the Main-Danube Canal commissioned by the German Federal Institute of Hydrology (BfG), Koblenz. This mission had the purpose to test the applicability of data of the High Resolution Stereo Camera – Airborne (HRSC-A; Hoffmann et al. 2000b; Wewel et al. 1998) for the classification of biotope types on reaches of Federal waterways characterized by strong relief features. The spatial resolutions amount to 30 cm in the case of the multi-spectral channels *(fig. 5)* and 200 cm in the case of the Digital Surface Model (DSM) which has bee derived by automatical matching (estimated accuracies 20 to 30 cm in planimetry, 50 cm in height). The DSM is resampled to 30 cm in order to use it in combination with the multispectral data set in the segmentation process. It is not possible to derive a reasonable normalized DSM because a Digital Terrain Model (DTM) is not available for the study site and an approximate normalization using appropriate methods yields unsatisfying results due to the sparse resolution and the limited reliability of the used elevation data (Schiewe 2001). A land use vector data set is available based on a visual interpretation of the HRSC-A data.

Aim of our study is the classification of the land use/land cover in this rural test site. The first step is the definition of a specific object catalogue. Due to the fact that there is no ground truth information available for the site the object catalogue is based on the visual interpretation of the data. The following objects define the object catalogue: channel, cultivated field, bare field, forest, smaller groups of trees, shrubbery and roads. The objects exhibit a large range of different scales ranging from small shrubs to large cultivated fields. Using the hierarchical approach of eCognition the segmentation is performed at different scale levels.

Due to the strictly sequential hierarchical approach problems with the inheritance of outlines of the scale level above can occur: Different input

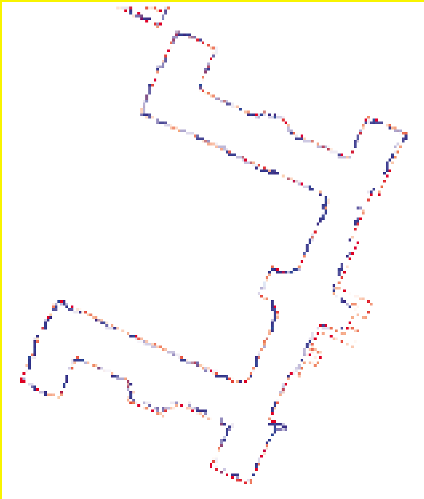


Fig. 4: Overlay of segmented building (Stuttgart Castle, subset: 200 x 250 m²; blue) onto cadastral data (red) showing the satisfying geometrical accuracy of the approach

sources may produce different segment outlines for the same object. Due to the heterogeneity of objects it is necessary to apply all input sources in a certain sequence in order to achieve good segmentation results for all objects. Here it can happen that segment outlines are inherited that do not fit to the objects that could be modelled best in the current level, so that either sliver polygons are generated, or even – if the statistical significance is not given any longer – no change at all occurs *(figure 6)*. In case of sliver polygons the sample set could become rather small so that the following classification could lead to unreliable results.

One solution to this problem is a manual post-processing. Here, eCognition supports interfaces for a user-defined and knowledge-based merge of neighbouring (sliver) segments. Because this is a rather labour-intensive approach and does not guarantee a perfect solution (in regions where no sliver polygons have been created), it should be possible to extract segments at any desired scale in a parallel manner and to fuse the (weighted) information not before the feature level. With this, the full potential of the input data could be utilized.

The best segmentation results are achieved with different spectral combinations and weighting factors for the different bands with respect to the object classes. *Figure 7* gives an example. The segmentation result shows clearly that it is not possible to detect the field with the only use of the panchromatic band *(figure 7, left)*, whereas the use of the red and near infrared band produces a good semantical separation *(figure 7, right)*. Here, the use of the DSM in the segmentation process shows no improvement which is due to the coarser resolution of the data and the lack of a normalized DSM.

4.4 Summary of results

With respect to the criteria mentioned above (robustness, grade of automatization and transparency) our empirical study on segmenting data of different test sites has revealed some promising but also several critical points for the general procedure as well as for the eCognition software in particular.

Concerning the robustness the general approach actually produces a significant progress compared to traditional methods: It is possible to extract homogeneous regions with unique semantical meanings and sufficient geometrical accuracy. It has to be pointed out that the term „multiscale“ processing used by the eCognition software is quite different from a conventional understanding: In fact, it is not possible to introduce original data of different spatial resolutions which leads to a considerable loss of information. Furthermore, the strictly sequential, hierarchical approach conflicts with the analysis of various multi-sensoral data combinations. Hence, it should be possible to extract segments at any desired scale and to fuse the information not before the feature level to make use of the full potential of the data sets.

With respect to the degree of automatization it has to be pointed out that important steps of the region-based method definitively imply time-consuming, manual interactions. Firstly, during the segmentation process the choice of scale parameters as well as the setting of weights for the input data sources have to be done by the user. Secondly, in contrast to human analysis the segmentation and the classification steps are strictly separated. Nevertheless, the segmentation has to produce a reasonable basis, so that the evaluation of the extracted regions implies a visual inspection. Thirdly, also the following fuzzification within the classification process is a costly process which is not automated.

Considering the transparency of the software in use it has to be stated that with increasing complexity the control and understanding significantly decrease. For instance, the effects of the abstract scale parameter settings can be hardly predicted. Here, the aim to combine various parameters to one number makes the initial use easier, but the actual, iterative process more difficult.

5 Conclusions

Traditional multi-spectral classification methods on pixel basis are no longer suited for the evaluation of high-resolution, multi-sensoral data from remote sensing. Region-based approaches consisting of a segmentation and a classification step have already proven to be a satisfying alternative solution.

Our empirical studies using multi-spectral imagery and elevation data have revealed some promising but also several critical points for the gen-

Fig. 5: Study area „Main-Danube Canal“ (channel combination near IR, green, blue; subset of 1100 x 900 m²)

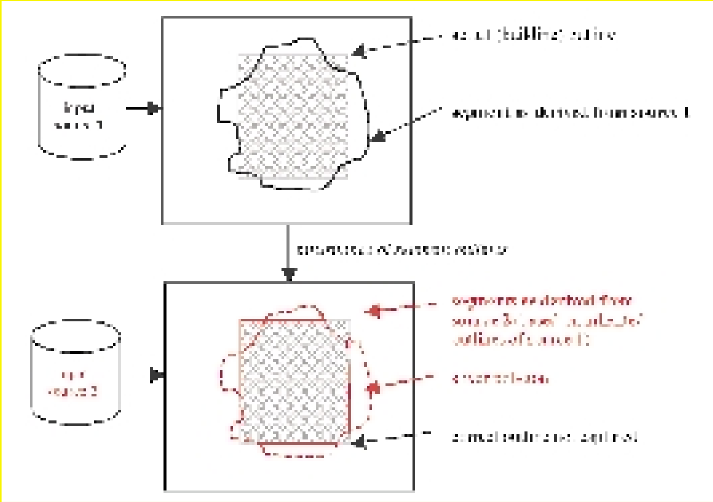


Fig. 6: Inheritance problems with sequential segmentation

eral procedure as well as for the applied eCognition software package in particular. The region-based approach, generating segments at various scales which are structured in a hierarchical manner, principally yields satisfying results with respect to the desired geometrical accuracy of outlines as well as to the unique class membership to one single region.

On the other hand, it can be observed that with increasing availability of multi-sensoral data a strictly sequential, hierarchical approach can lead to undesired sliver polygons or even missing segments. Also an introduction of actual multi-scale data is not possible. Hence, a segmentation should also allow a parallel segment generation and an information fusion after the feature level.

Like other traditional approaches the presented region-based method still incorporates a couple of time-consuming manual and visual interactions. In this context, from our point of view more efforts should be placed on an integration of model-based procedures even prior to the segmentation step to define such attributes and relations that produce characteristic homogeneities within one class and heterogeneities to neighbouring classes at specific scales.

Finally, it can be stated that the presented methods and implementations are representing an important progress because it is obvious that a multisensoral analysis is needed to realize the full potential of remotely sensed data. This trend will even be strengthened by the sensors that will simultaneously acquire different data types (like electro-optical imagery and laserscanning data) which are announced for the near future.

6 Acknowledgements

The laserscanning and the cadastral data for the test site „Stuttgart“ is courtesy of the OEEPE-project on „Laser Data Acquisition“. The digital orthoimage was provided through the State Survey of Baden-Württemberg (www.lv-bw.de/LVShop2/ index.asp). The data set „Main-Danube Canal“ is courtesy of the German Federal Institute of Hydrology (BfG), Koblenz.

7 References

APLIN, P., ATKINSON, P. M. AND CURRAN, P. J. (1999): Per-field classification of land use using the forthcoming very fine spatial resolution satellite sensors: problems and potential solutions. In: Atkinson, P.M. and Tate, N.J. (Eds.): Advances in remote sensing and GIS Analysis. Chichester, John Wiley & Sons: 219–239.

ARBOREAL (2001): www.arboreal.fi

BAATZ, M. AND SCHÄPE, A. (2000): Multiresolutional segmentation – an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T. and Griesebner (Eds.): Angewandte Geographische Informationsverarbeitung XII: 12–23.

CUSHNIE, J. L. (1987): The interactive effect of spatial resolution and degree of internal variability within land-cover types on classification accuracies. International Journal of Remote Sensing. 8(1): 15–29.

DEFINIENS (2001): www.definiens.com

EBNER, H., BAUMGARTNER, A. AND STEGER, C. (1998): Automatische Extraktion von Straßen aus Luftbildern. Österr. Zeitschrift für Vermessung und Geoinformation. (85)3: 177–186.

FABER, A. AND FÖRSTNER, W. (1999): Scale Characteristics of Local Autocovariances for Texture Segmentation. International Archives of Photogrammetrie and Remote Sensing. 32(7–4–3), Valladolid, Spain.

HANDELS, H. (2000): Medizinische Bildverarbeitung. Stuttgart, Teubner.

Handmann, U., Lorenz, G., and von Seelen, W. (1998): Fusion von Basisalgorithmen zur Segmentierung von Straßenverkehrsszenen. Mustererkennung. Heidelberg, Springer Verlag: 101–108.

HARALICK, R. AND SHAPIRO, L.G. (1985): Image segmentation techniques. Computer Vision, Graphics and Image Processing 12: 100–132.

HOFFMANN, A., SMITH, G. M. AND LEHMANN, F. (2000a): Die Klassifizierung hochauflösender Daten – ein Per-Parcel-Ansatz mit Daten des digitalen Kamerasystems HRSC-A. In: Strobl, S., Blaschke, T. and Griesebner, G. (Eds.): Angewandte Geographische Informationsverarbeitung XII.. Heidelberg, Wichmann: 228–237.

HOFFMANN, A., VAN DER VEGT, J. W. AND LEHMANN, F. (2000b): Die automatische Aktualisierung von Karten: Neue Methoden mit Daten der digitalen Kamera HRSC-A. Kartographische Nachrichten 50(5): 211–217.

JÄHNE, B. (1997): Digitale Bildverarbeitung. Berlin, Springer.

RYHERD, S. AND WOODCOCK, C. (1996): Combining spectral and texture data in the segmentation of remotely sensed images. Photogrammetric Engineering & Remote Sensing 62(2): 181–194.

SCHIEWE, J. (2001): Ein regionen-basiertes Verfahren zur Extraktion der Geländeoberfläche aus Digitalen Oberflächen-Modellen. Photogrammetrie – Fernerkundung – Geoinformation. 2: 81–90 (in print).

TOPOSYS (2001): www.toposys.com

WESFREID, E. AND WICKERHAUSER, V. (1999): Vocal command signal segmentation and phonems classification. Wireless Networks.

WEWEL, F., SCHOLTEN, F. NEUKUM, G. AND ALBERTZ, J. (1998): Digitale Luftbildaufnahmen mit der HRSC – Ein Schritt in die Zukunft der Photogrammetrie. Photogrammetrie – Fernerkundung – Geoinformation 6: 337–348.

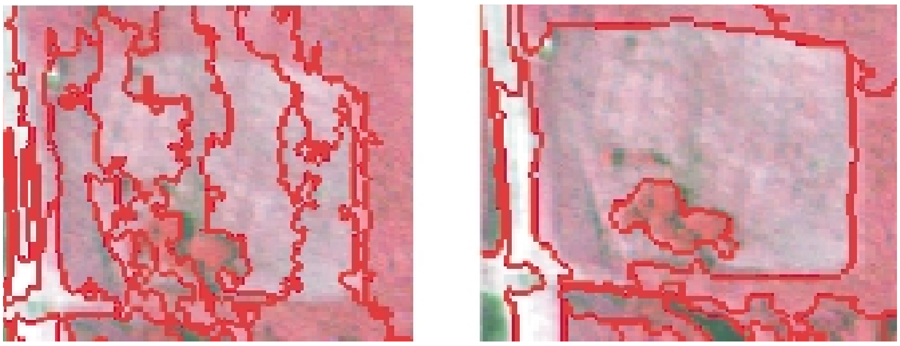


Fig. 7: Superior segmentation results using multi-spectral input (right) compared to pan-chromatic data (left)