FUSION OF COLOUR INFRARED IMAGERY AND AIRBORNE LASER SCANNING DATA IN THE AUTOMATIC CLASSIFICATION OF URBAN ENVIRONMENTS

[1]J. B. K. Kiema
[2]H.-P. Bähr

Institute for Photogrammetry and Remote Sensing (IPF)
University of Karlsruhe
Englerstr. 7, 76128 Karlsruhe, Germany
Tel.: +49-(0)721-608-6994, Fax: +49-(0)721-694-568
[1]kiema@ipf.uni-karlsruhe.de
[2]baehr@ipf.uni-karlsruhe.de

ABSTRACT

The fusion of multi-sensor data is important in the automatic classification of urban environments. In this study, the additional channel concept is used to fuse airborne laser scanning (ALS) data with a colour infrared (CIR) image acquired for a typical urban scene. To facilitate the co-registration of the two data sets is generated a CIR ortho-image using the digital surface model derived from the ALS data. This effectively integrates the geometric and multi-spectral information sources. In order to incorporate context analysis in the feature extraction, the feature base is expanded to include both spectral and spatial features. A maximum likelihood classification approach is then applied after an initial clustering procedure. Shadow effects present in the fused imagery are subsequently eliminated. The error matrix method is then employed to assess the thematic accuracy of the classification results obtained. Kappa related measures as well as accuracy parameters derived from Kullback-Leibler information on multinormal distributions are estimated. For comparison purposes, use is made of Delaunay triangulation. It is demonstrated that the classification of urban scenes is significantly improved by fusing the multi-spectral and geometric data sets.

Key words: Data fusion, airborne laser-scanning, multisource classification .

1. INTRODUCTION

The future use of earth sensor data and through this, the future development of mapping sensors, will continue to be influenced by algorithmic progress in various fields including: automatic image understanding, data fusion and data compression. In this regard, the automatic classification of features from remotely sensed imagery is important for GIS database updating. This is of particular interest in urban environments, especially given the high concentration of man-made features in such areas. Indeed, that most of the research effort in the automatic segmentation of geo-spatial imagery is currently focused on man-made features in general, and buildings and roads in particular, testifies to this fact. Nonetheless, in view of their size, structure, spectral characteristics as well as simple diversity, the segmentation of man-made features poses a special challenge.

Spectral information is conventionally employed in the classification of multi-spectral imagery. However, results obtained from such methods are usually unsatisfactory, particularly for applications involving the mapping of man-made structures and natural features in complex urban scenes. These are often characterised by limited accuracy and low reliability (Haala and Brenner, 1999). Furthermore, it is not possible to discriminate between object features that display similar spectral reflectance characteristics for instance, building roofs from pavements that are built using similar materials or trees from grass-covered areas.

Airborne laser scanning (ALS) is well suited for the production of digital surface models (DSMs). The geometric information contained in this data can be used to support the segmentation of objects that are projected higher than the terrain (e.g., buildings, trees etc.) from those that are basically at terrain level (e.g., pavements, gardens, parks etc.) (Haala and Walter, 1999). The use of this data alone is however, of limited applicability, especially in the segmentation of urban objects. This is essentially because of the restriction of surface geometry which limits the number of object types that can be discriminated within the DSM. In order to enhance its usefulness and thereby exploit its full potential, it is often necessary to integrate this with other complementary data sources e.g., multi-spectral data.

From the foregoing, the main objectives of the study presented here are twofold. Firstly, to consider within the perspective of data fusion, the automatic classification of urban environments from colour infrared (CIR) imagery and ALS data. And secondly, to compare this with similar results obtained from Delaunay triangulation methods.

2. TEST AREA AND DATA USED

The selected test area for this study encompasses a typical urban scene in the city of Karlsruhe. This lies in the south-western part of the Federal Republic of Germany near the boundary with France and includes part of the main campus of the University of Karlsruhe. CIR imagery captured at a scale of 1:5 000 using a conventional aerial camera are scanned at $170\mu m$. A normalised DSM derived from ALS data is also employed. An ortho-image is generated from the CIR image by using the available DSM. This enables the co-registration of the different data sets while at the same time correcting for the relief displacement in the CIR image. Fig. 1 shows the generated CIR ortho-image.

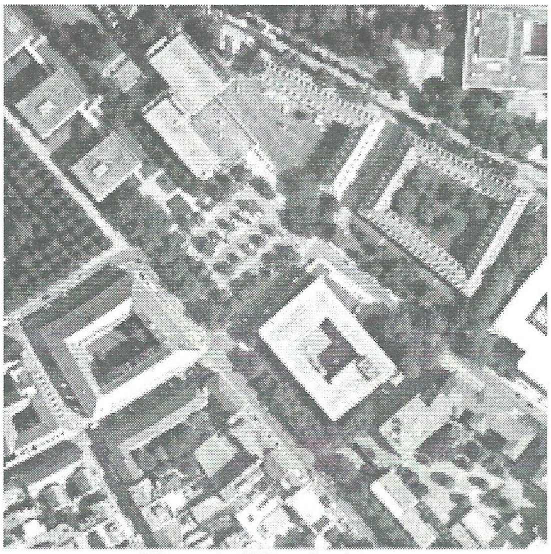

Figure 1: CIR ortho-image

In principle, two ALS sensor types can be distinguished: pulse and continuous wave (CW) laser systems. The operating principle of ALS is relatively simple. Basically, a laser beam is emitted from a sensor usually on-board an airborne platform. The signal as reflected from different objects or the terrain surface underneath is then recorded. For the pulse laser system, the range distance from the sensor to the reflecting objects/surface is estimated from the run-time between emission and reception of this signal. The same is estimated for the CW laser system from phase difference measurements as articulated in (Wehr and Lohr, 1999).

Using this range distance together with the instantaneous sensor position and orientation obtained from integrated GPS/INS, it is possible to estimate the position and elevation of different object features. Object points can be measured for every $0.5x0.5m^2$ with accuracies of the order of about $0.3m$ (Lohr, 1997). Fig. 2 exemplifies a typical 3D view of a DSM obtained from the first pulse (echo) measurement for part of the ALS data used in this study.

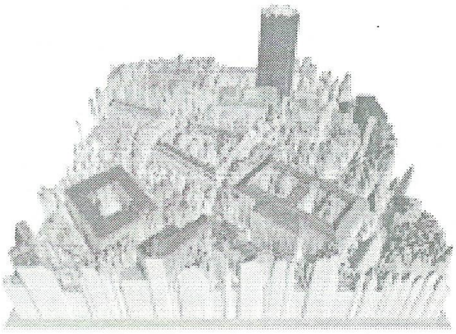

Figure 2: 3D illustration of a DSM derived from ALS data

3. DATA FUSION

3.1. Concept of data fusion

The quantity of geo-spatial data available for the study of the environment, the geosphere and the biosphere has grown significantly over the last couple of years, partly due to the increase in the number of space missions devoted to the observation of the earth. Faced with the scenario of an ever increasing amount of data, it is imperative that the best information for a particular application be extracted. In this regard, and in order to facilitate an optimal exploitation of the synergy of these data, data fusion techniques are of paramount importance. (Wald, 1998) defines data fusion as a formal framework in which are expressed means and tools for the alliance of data originating from different sources. Conceptually, this aims at obtaining information of greater quality. The phrase "greater quality" is used here in a generic sense and hence, its exact meaning will vary depending on the application under consideration.

Although the concept of data fusion is easy to understand, its exact meaning and use often varies from one scientist to another. In principle, this may be performed at different levels for example, at measurement, attribute, rule or even decision levels.

Different approaches to the fusion of geo-spatial data may theoretically be employed e.g., RGB colour composites, IHS transformation, Principal Component Substitution, wavelets etc. (Pohl, 1999). The particular method adopted is influenced by several factors including: the type of application under study, the structure of the data to be fused as well as the image characteristics that need to be enhanced or preserved.

3.2. Fusing the different data sources

The fundamental ideas behind the combination of multisource data for scene labelling are outlined in (Hahn and Stätter, 1998). Similarly, the combination of image and range data in the reconstruction of buildings is discussed in (Haala, 1996). In the application considered in this study, the geometric information of the ALS data is fused with the multi-spectral information of the CIR. In this regard, several different approaches to data fusion can be adopted. For instance, it is possible to make use of the hierarchical classification approach, the additional channel concept or even, a knowledge-based strategy

The hierarchical or layered classification approach is basically a structured technique through which the different data sets to be fused are applied in such a way as to successively divide the working area into more detailed object classes (Savian and Landgrebe, 1991). In principle, this begins with basic object classes before progressively zeroing in on more detailed ones. On the other hand, in the additional channel method the different data sets are introduced as separate channels in an integrated fashion within an expanded data framework. This of course necessitates the co-registration of the different data sets to a uniform georeference system. For the knowledge-based procedure, the data to be fused needs to be appropriately modelled and structured with respect to the particular knowledge representation formalism adopted (e.g., rule-based systems, semantic networks etc.). An example of the use of semantic networks in the interpretation of digital aerial photographs using a map-supported approach is discussed in (Quint, 1997).

Comparing the above methods, the main disadvantage of the hierarchical classification approach is the propagation of the classification errors in the subsequent classification steps. Conversely, the main advantage of the additional channel procedure is its simplicity as well as the enhanced flexibility in the data processing. On the other hand, although knowledge-based methods for the integration of multi-source data are fairly rigorous, these are nevertheless, relatively complicated as the entire data system needs to be reorganised accordingly. This can be a fairly complex task depending on the intricacy of the particular problem under investigation and the structure of the different data sets that need to be integrated. Because of the above reasons, the additional channel concept is adopted for the data fusion in this study. Hence, the normalised ALS data is integrated as an additional channel alongside the three multi-spectral bands of the generated CIR ortho-image.

3.3 Limitations in the fused data

Experience with the fused data has shown that it is desirable that the different data sets to be synergised be captured in more-or-less the same season, particularly if multi-spectral image data acquired from satellite sensors with sun synchronous orbits is employed. This is important in the selection of the training data as it avoids the imperative need to digitise new training areas with every different data set integrated. Normally, because of different spectral characteristics of different object features (e.g., vegetation) depending on the season of data acquisition, different spectral diffusions are often recorded. To overcome this problem, new training areas need to be ideally digitised for every fused data set. In practice, this translates to a very time consuming exercise. A more practical solution to this would be to procure the training data using automatic methods from an already existing GIS database as proposed in (Walter, 1998).

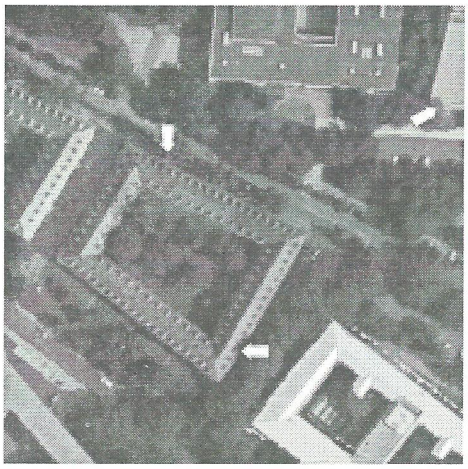

Figure 3: Occlusions near buildings affect the homogenity of multi-spectral data

Besides the above drawback, another open issue remains the homogeneous quality of the different data sources employed. In order to fully exploit the integrated data, the homogeneity of all the fused data sets should ideally be maintained for the entire test area. In practice however, because of several reasons this is not always upheld. Inevitably, this leads to problems in the image segmentation. Several possible manifestations of this are noted. Firstly, as shown in Fig. 3 the homogeneity in the case of high-resolution multi-spectral data may be infringed by occlusions near buildings often as a result of nearby trees. This is particularly complicated in urban scenes where it is ordinarily difficult to distinguish between buildings and trees using their spectral characteristics, especially if both exhibit similar height and texture. As discussed in (Centeno *et al.*, 1999), a possible solution to this lies in the introduction of threshold height values to the segmented candidates for the class *Building*. The obstruction of object features resulting purely from shadows which incidentally also affects the homogeneity of fused data, especially for high-resolution geo-spatial image data, is examined here separately from a post-processing perspective.

Geometric problems in the generated CIR ortho-image may also compromise the homogeneity of the fused data. Similarly, limitation in the ALS data may lead to certain areas in the object space being omitted. This manifests itself as gaps or "dead areas“ within the DSM. Fig. 4 shows an example of this problem for the first laser pulse measurement with the dark regions denoting the "dead areas“. In principle, these undesirable regions are the result of poor or no object reflection being recorded from those particular areas at the adopted wavelength of the laser beam. Certainly, this reiterates the importance of evaluating the backscattering properties of the object features of interest when considering the appropriate laser wavelength for a particular task (Wehr and Lohr,1999). In general, specular surfaces (e.g., water bodies) often result in low reflectance values.

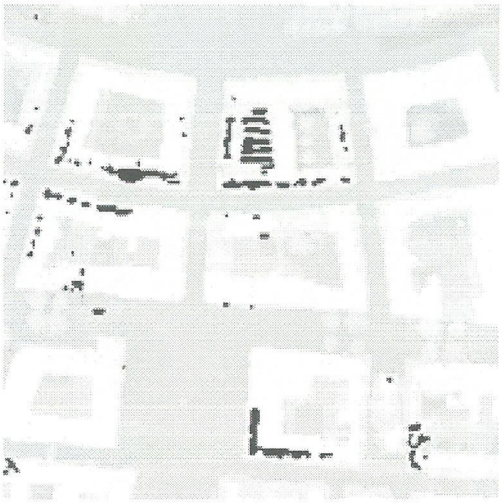

Figure 4: ALS data showing "dead areas“

4. CLASSIFICATION OF URBAN ENVIRONMENTS

4.1. Integrated classification approach

The inadequancy of conventional multi-spectral classification in urban areas has been acknowledged in many studies (e.g., (Fung and Chan, 1994); (Barnsley and Barr, 1996)); (Haala and Brenner, 1999) etc). In applications where high accuracies are required, it is imperative to enhance the feature base. This needs to be expanded to include both spectral and spatial parameters ((Schilling and Vögtle, 1996); (Kunz *et al.*, 1998)). Table 1 shows some of the parameters that need to be considered. These may be included either explicitly through the introduction of complementary data sources or implicitly through the use of appropriate segmentation methods. For example, it is possible to implicitly incorporate topological information through the use of triangulation segmentation methods e.g., Delaunay triangulation.

Spectral features	Spectral Signature Texture
Spatial features	Structure Size Shape/Contour Topology

Table 1: Extended feature base

Different classifiers may be employed in the supervised classification of remotely sensed data, namely: maximum likelihood, minimum distance and parallelepiped classifiers (Lillesand and Kiefer, 1994). For the study presented here, a maximum likelihood classification approach, as opposed to contextual segmentation methods, is adopted after an initial clustering procedure. The selection of the training data is basically done using manual digitising. Five basic urban object classes are identified: *Buildings, Pavements, Trees, Grass-covered areas* and *Special.* The class *Special* is introduced in order to take care of the many miscellaneous urban objects of limited dimension (e.g., vehicles, water fountains, sculptures etc.) that otherwise exhibit a disturbing influence on the segmentation of other object classes. In order to circumvent this problem, these features are treated as a separate object category and their unique characteristics (e.g., size) exploited to enable their segmentation. Moreover, it is important to handle these objects separately since most of these are basically "non-permanent features" which can later on be re-classified into the more permanent object features that they obscure (e.g., pavements). The classification results obtained when only the CIR image is employed are shown in Fig. 5.

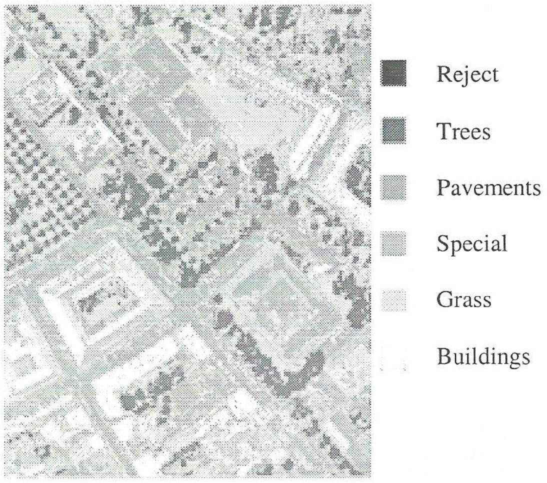

Figure 5: Classification results for multi-spectral data excluding ALS data

Prior to the final classification, the elimination of shadow effects is considered. In general, shadows constitute a major impediment to the successful segmentation of areas that are characterised by rough surfaces and steep slopes, as is often the case in urban scenes. Basically, these "falsify" the spectral response of the features that they obscure thereby distorting the overall image homogeneity. This effect is more apparent in high-resolution remotely sensed imagery than in corresponding low-resolution imagery. The elimination of shadow effects is therefore most desirable. For the purposes of this study, this is achieved using an approach similar to that described in (Haala and Walter, 1999).

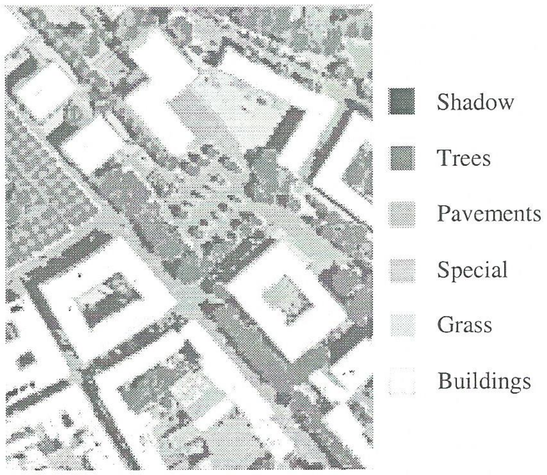

Figure 6: Classification results for fused CIR ortho-image and ALS data showing shadow effect

Firstly, the shadow areas on the imagery are automatically identified. A separate class *Shadow* is then introduced for each of the apriori defined object classes. This means that each object class is divided into one separate subclass for shadow areas and another for non-shadow areas. Separate training data is digitised for each subclass. The classification is then carried out after which the shadow and non-shadow subclasses for each object class are combined. This results in one unique class for each apriori defined object class. Fig. 6 illustrates the classification results when predefined shadows are introduced as a separate class. The final classification results in the presence of both the multi-spectral and geometric data are shown in Fig. 7.

Normally, after classifying an image some quality measure is required in order to allow a degree of confidence to be attached to the results. This also serves to indicate whether the analysis objectives have been realised (Richards, 1993). Different quality aspects may be evaluated depending on the particular analysis domain considered for instance, accuracy, resolution, completeness, consistency etc. This study focuses on the thematic accuracy of the classification results. In general, different approaches to the assessment of this can be distinguished including: error matrix-based methods, spectral distance-based methods as well as quantitative methods.

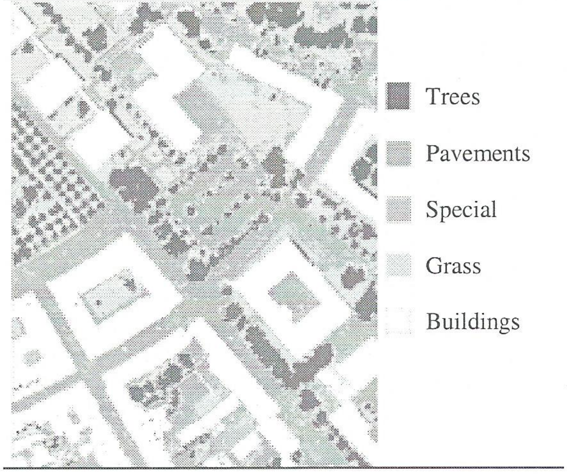

Figure 7: Classification results for fused CIR ortho-image and ALS data

The error matrix method is used to evaluate the thematic accuracy of the final classification results obtained. Ground truth methods complemented by existing topographic map sheets are employed as the reference basis for the evaluation. Stratified random sampling procedures are used to estimate the error (confusion) matrix. Various Kappa related measures as well as accuracy parameters derived from Kullback-Leibler information on multinormal distributions (Nishii, 1999) are then estimated as presented in Table 2. These results confirm that a fairly good image classification is realised.

Kappa statistic:	0.902
Class-averaged accuracy:	0.908
Overall accuracy:	0.902
Jpro:	0.901
Juni:	0.908
R_pro$(X \mid \Gamma)$:	0.098

Table 2: Thematic classification accuracy

4.2. Delaunay Triangulation

Triangulation-based segmentation methods are generally employed in mid-level image processing procedures in order to combine structured image regions into semantically homogeneous clusters. The fundamental image segmentation primitive in triangulation schemes is the *Delaunay triangle* (Centeno and Weindorf, 1999). Through this, use is made of the topological relationship between the image segments. This is in contrast to conventional segmentation methods which rely on the spectral information of the image pixels. In principle, the image is first segmented using some low-level image analysis procedure. Thereafter, a Delaunay tessellation is developed from the extracted image segments. This in effect transforms the image pixels into the graph structure (Anders, 1999).

In general, a triangulation is defined as a subdivision of an area into triangle primitives. The Delaunay triangulation (DT) which is essentially a graph-based clustering method, has the intrinsic property that the circumcircles of every triangle are empty (Okabe *et al.*, 1992). The result of this is a set of polygons describing the semantically homogeneous region of the image that have a unique topological structure. In order to smoothen the extracted segments and minimise the level of noise in the segmentation results, use is made of mathematical morphology and connected components. A more comprehensive discussion of these is given in (Serra, 1986) and (Haralick and Shapiro 1992) respectively. As an example, segments are extracted for the object class *Buildings*. A Delaunay tessellation is then applied. Fusing the valid object segments as described in (Schilling and Vögtle, 1996) results in Fig. 8.

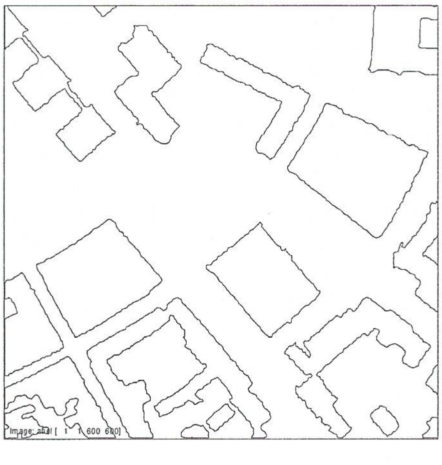

Figure 8: Selection and fusion of valid building segments

4.3. Discussion and Application of Results

A number of basic observations can be drawn from the above results. Firstly, it is virtually impossible to effectively classify CIR imagery covering urban scenes when only the spectral information is relied upon. A comparison of Figs. 5 and 7 clearly demonstrates the improvement realised in the classification results upon fusing the multi-spectral and geometric data sets. In particular, the enhanced ability to discriminate between low lying urban objects (e.g., pavements) from objects that are significantly above the terrain (e.g., buildings) is noted. In addition, the ability to distinguish pavements from buildings that are characterised by roofing material that exhibit similar spectral characteristics to pavements is also observed. The possible misclassifications that would result in the event that shadow effects are not adequately corrected for is demonstrated by comparing Figs. 6 and 7. This reaffirms the importance of correcting for shadow effects in high-resolution remotely sensed imagery. In general, the Delaunay triangulation results compare very well with similar results obtained using the integrated classification approach, particularly in as far as the building contours are concerned.

Although the classification results obtained are neither precise nor detailed enough to enable the automatic reconstruction of buildings or roads they can nevertheless, be employed in the automatic verification of objects defined in GIS databases. Through this, the spatial inconsistencies between the existing GIS database and the classified image can be automatically assessed. This can then be used to assist an operator-based updating of the GIS data. Another possible application of the classification results would be in the mapping of urban vegetation (e.g., tree coverage). Extracted trees could be used to support visualisation and simulation of urban environments using 3D city models or in forest management. In addition, these could also be used to provide context information in the automatic extraction of other urban objects as noted in (Baumgartner, 1998). Similarly, the Delaunay triangulation results may be used to support the automatic verification of GIS databases.

5. <u>SUMMARY AND CONCLUSSIONS</u>

This study underlines the importance of data fusion in the classification of complex urban scenes. In particular, the need to fuse multi-spectral and geometric data sets is underscored. This provides supplementary object information that facilitates the discrimination of the various man-made and natural objects. In this regard, it can be projected that, despite the high anticipation associated with the introduction of high-resolution commercial sensors, the importance of data fusion in the classification of urban environments will probably remain. Nonetheless, it is imperative that the fusion of multi-source data be formulated and structured accordingly in order to enable users to fully rely on software to integrate their data.

The need to incorporate context information in the feature extraction is also articulated. This is achieved by expanding the feature base to incorporate both spectral (e.g., spectral signature, texture) and spatial features (e.g., size, structure, topology etc.). Further, it is postulated that as the spatial resolution of satellite sensor imagery continues to increase, the need to introduce local context analysis in the automatic recognition of urban objects is going to become even more necessary.

6. ACKNOWLEDGEMENT

The authors wish to acknowledge the support of the *Deutscher Akademischer Austauschdienst* (DAAD) in the preparation of this paper.

7. REFERENCES

ANDERS, K.-H., Sester, M. and Fritsch, D., 1999: Analysis of Settlement Structures by Graph-based Clustering, In: *Semantic Modelling for the Acquisition of Topographic Information from Images and Maps*, pp. 41-49.

BARNSLEY, M.J., and Barr, S.L., 1996: Inferring Urban Land Use from Satellite Sensor Images using Kernel-based Spatial Reclassification, *Photogrammetric Engineering and Remote Sensing*, Vol. 57(4), pp. 949-958.

BAUMGARTNER, A., 1998: Extraction of Roads from Aerial Imagery based on grouping and local context, In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 3/1, pp. 196-201.

CENTENO, J.A.S., Kishi, R.T. and Bähr, H.-P., 1999: Recognition of Buildings Using Scanned Maps and Laser Scanner Altitude Data, *Photogrammetrie. Fernerkundung. Geoinformation.* Vol. 1/99, pp. 19-28.

CENTENO, J.A.S., and Weindorf, M., 1999: Spatial Information for Image Segmentation, XVII CIPA Symposium, 3-6 October, 1999, Olinda, Brasil.

FUNG, T. and Chan, K.C., 1994: Spatial Composition of Spectral Classes: A Structural Approach for Image Analysis of Heterogeneous Land-use and Land-cover types, *Photogrammetric Engineering and Remote Sensing,* Vol. 60(2), pp. 173-180.

HAALA, N., 1996: Gebäuderekonstruktion durch Kombination von Bild- und Höhendaten. Dissertation, Deutsche Geodätische Kommission, Reihe C, Heft Nr. 460, München.

HAALA, N. and Brenner, C., 1999: Extraction of Buildings and Trees in Urban Environments, *Photogrammetric Engineering and Remote Sensing,* Vol. 54(2-3), pp. 130-137.

HAALA, N. and Walter, V., 1999: Automatic Classification of Urban Environments for Database Revision using Lidar and Colour Aerial Imagery, In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 3/1, pp. 76-82.

HAHN, M. and Stätter, C., 1998: A Scene Labelling Strategy for Terrain Feature Extraction using Multisource Data, In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 3/1, pp. 435-441.

HARALICK, R. and Shapiro, L., 1992: Computer and Robot Vision, Vol. 1, Addison-Wesley, Reading Massachusetts.

KUNZ, D., Vögtle, T. and K.-J Schilling., 1998: Integrierte Verarbeitung von Satellitenbild- und vektorieller Karteninformation. In : *Digitale Bildverarbeitung*, Bähr, H.-P. and Vögtle, T. (eds.), Wichmann Verlag, Karlsruhe.

LILLESAND, T. and Kiefer, R., 1994: Remote Sensing and Image Interpretation, John Wiley and Sons, New York.

LOHR, U., 1997: Digital Elevation Models by Laserscanning: Principle and Applications. In *Third International Airborne Remote Sensing Conference and Exhibition*, pp.174-180.

NISHII, R. and Tanaka, S., 1999: Accuracy and Inaccuracy Assessment in Land-cover Classification. In *IEEE Transactions on Geoscience and Remote Sensing*, Vol 37(1), pp. 491-498.

OKABE, A., Boots, B. and Suguhara, K., 1992: Spatial Tesselations: Concepts and Application of Voronoi Diagrams. John Wiley and Sons, Chichester.

POHL, C., 1999: Tools and Methods for Fusion of Images of Different Spatial Resolution, In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 7-4-3, pp. 76-82.

QUINT, 1997: Kartengestützte Interpretation monokuler Luftbilder, Dissertation, Deutsche Geodätische Kommission, Reihe C, Heft Nr. 477, München.

RICHARDS, B., 1993: Remote Sensing and Digital Image Analysis. Springer-Verlag, Berlin.

SAVIAN, F. and Landgrebe, D., 1991: A Survey of Decision Tree Classifier Methodology, *IEEE Transactions on Systems, Man and Cybernetics 21*, No. 3, pp. 660-674.

SCHILLING, K-J. and Vögtle, T., 1996: Satellite Image Analysis Using Integrated Knowledge Processing, In *International Archives of Photogrammetry and Remote Sensing, XVIII Congress of the ISPRS*, Part B3, Commission III, pp. 752-757.

SERRA, J., 1986: An Introduction to Mathematical Morphology, *Computer Vision, Graphics and Image Processing*, Vol. 35, pp. 283-305.

WALD, L., 1998: A European Proposal for Terms of Reference in Data Fusion, In *International*

Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 7, pp. 651-654.

WALTER, V., 1998: Automatic Classification of Remote Sensing Data for GIS Database Revision, In *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 4, Commission IV, pp. 641-648.

WEHR, A. and Lohr, U., 1999: Airborne Laser Scanning: An Introduction and Overview. *Photogrammetric Engineering and Remote Sensing,* Vol. 54