FUSION AND PERCEPTUAL ORGANISATION OF FEATURES FROM MULTI-SENSOR DATA: GENERAL CONCEPTS AND NEW DEVELOPMENTS

J. Schiewe

University of Vechta, Research Center for Geoinformatics and Remote Sensing (FZG), PO Box 1553, 49364 Vechta, Germany – jschiewe@fzg.uni-vechta.de

Commission IV, WG IV/7

KEY WORDS: fusion, interpretation, recognition, feature, laser scanning, multi-sensor

ABSTRACT:

While the available, eventually multi-source data from remote sensing systems yield the advantage of producing more features, the crucial point in the scene interpretation workflow is still the big, application dependent gap between these features and the related object characteristics. The necessary bridging process of perceptual organisation as well as the fusion of features have been neglected too much in the past. In this contribution we will present our methodological developments concentrating on the perceptual organisation of features that have been derived from multi-sensor data aiming at the extraction and classification of topographic surface edges. The algorithm we are proposing consists of a segmentation and classification of the raster-based elevation and image data, transforming the resulting polygon outlines into vector format and performing some post-processing steps (dilation, smoothing).

1. INTRODUCTION

The various and heterogeneous user demands concerning properties of remotely sensed data have led to the development of a couple of new sensors with advanced spatial, spectral, radiometric or temporal characteristics. Furthermore **multi-sensor systems** have been designed for the simultaneous acquisition of image and elevation data, in particular by electro-optical cameras and laser scanning systems. Section 2 gives an overview of such multi-sensor data sources in general, and the FALCON system, which will be used in this study, in particular.

However, despite these technical developments the user's acceptance for remote sensing products has not been significantly increased. This is mainly due to the fact that respective **automatic processing methods** – in particular for interpretation purposes – are not operational or have not revealed an additional value of the advanced data sources yet. To be more specific: While the available, eventually multi-sensor data yield the advantage of producing more scene features, the crucial point in the interpretation workflow is still the big, application dependent gap between these features and the related object characteristics (Schenk, 2003). The necessary bridging process of perceptual organisation has been neglected too much in the past. Furthermore, a fusion of features coming from different sensors has hardly been applied within the perception step. Section 3 will elaborate on the aspects of perceptual organisation and fusion in more detail from a theoretical point of view.

In the context of these deficiencies we will demonstrate in the second, practical part of this contribution our methodological developments concentrating on a fusion of features that have been derived from multi-sensor data. As an example we will focus on the **extraction of topographic surface edges** (commonly known as breaklines). The task from a methodological point of view consists not only in delineating the edges as such but also in their classification (e.g. walls, embankments). The algorithm we are proposing consists of a segmentation and classification of both, the raster-based elevation and image data, transforming the resulting polygon outlines into vector format and performing some post-processing steps (dilation through matching with image edges and smoothing). This hybrid feature level fusion process as well as first qualitative results will be presented in section 4.

2. MULTI-SENSOR DATA SOURCES

2.1 General remarks

Due to the enhanced performance of GPS/IMU (Inertial Measurement Units) solutions for capturing position and orientation data of the associated moving platforms airborne laser scanning systems have reached maturity in the last ten years. Today standard systems capture both, multiple reflections (first and last echo data) and intensities of the reflections (generally, in the near infrared portion of the spectrum).

Enabling furthermore the simultaneous acquisition of multi-spectral image data, one yields high potential *multi-sensor systems* that on one hand deliver more accurate and reliable elevation data compared to stereo image matching solutions, and on the other hand by-pass the disadvantage of the "blind" laser scanning information. The presently most prominent examples of these multi-sensor airborne systems are FALCON (TopoSys company), which will be used within this study (see section 2.2), the laser scanner ALS 50 in combination with the

ADS 40 camera (Leica Geosystems) or the ALTM-3033/3070 (Optech) with the Emerge DSS camera. Schiewe (2004) gives a detailed overview of these systems.

2.2 FALCON sensor system

In 2002 the company TopoSys (Germany) released its airborne FALCON system which not only delivers laser scanning elevations but also acquires multi-spectral imagery simultaneously (Lohr, 2003). In the following a scene covering the City Memmingen (located in Southern Germany) and its neighbourhood will be processed.

The **imaging sensor** consisting of a line array scanner that acquires data in four bands in the visible and near infrared spectrum with a ground pixel size of 0.5 m and at a radiometric resolution of 11 bit.

The **laser scanner** operates with a glass fibre array producing a parallel acquisition pattern. The resulting elevation data with a height accuracy of about ± 0.2 m are delivered with a raster grid size of $(0.5 m)^2$ or $(1.0 m)^2$, respectively, having a swath width of 250 m for a flying altitude of 1000 m.

A special feature which is of major interest for the following study is the ability to capture *multiple reflections* of the laser beam (Löffler, 2003): From each laser pulse of the FALCON system the respective lowest and highest values (first echo and last echo, respectively) are recorded (other systems might record even more reflections). It is important to note that this separation is only possible if the two echoes are at least one pulse length (i.e. 5 ns, corresponding to 1.5 m, taking the speed of light into account) apart from each other. That means, that rather low objects like bushes cannot be separated from the ground by subtracting first and last echo. On average one
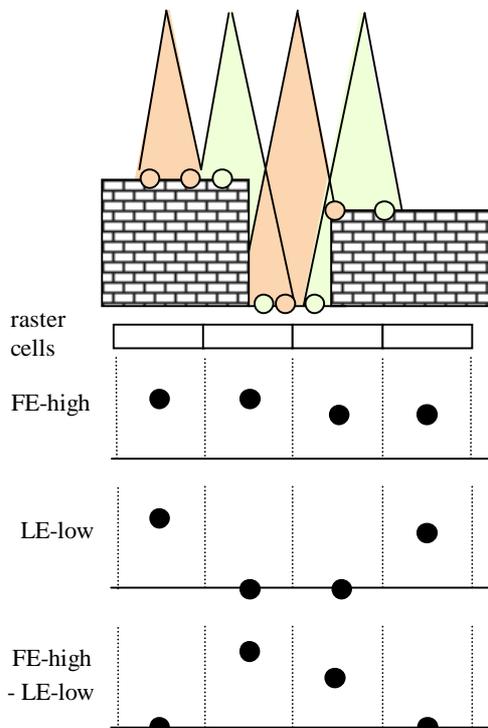

Figure 1. Schematic profile representation of a building through different reflections

grid cell of $(1.0 m)^2$ of the derived Digital Elevation Model is "hit" by a couple of laser beams. Hence, we have the possibility of distinguishing the respective highest and lowest values of the first and last echoes, respectively (called FE-high, FE-low, LE-high, LE-low). As a consequence, the respective elevation models differ in their representation of the scanned surface. For example (see figures 1 and 2), the first echo-highest values (FE-high) are able to represent narrow, outstanding objects (like buildings and vegetation) while narrow gaps in between disappear. Outstanding objects are extended in size. On the other hand, the last echo-lowest values (LE-low) show the opposite properties. Consequently, the difference (FE-high minus LE-low) leads to the representation of vegetation and (rather broad) building edges.
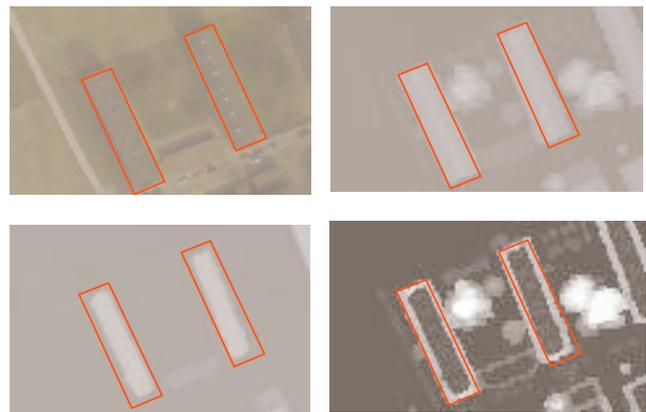

Figure 2. Different representation of buildings through different laser pulse reflections: image data (upper left, for comparison purposes), FE-high (upper right, buildings too large), LE-low (lower left, buildings too small) and difference FE-high minus LE-low (lower right, broad building edges).

3. GENERAL REMARKS ON PERCEPTUAL ORGANISATION AND FUSION

The process of (automatically) interpreting a remotely sensed scene is strongly correlated to the process of *cognitive perception* (see also figure 3). The latter does not only include the acquisition and representation of various stimuli by a human being, but also their organisation and interpretation. Cognitive perception can be seen as a hybrid process, i.e. it contains procedures in both directions, top-down (or model driven) and bottom-up (or data driven) at the same time. Furthermore, it is not simply a linear process, but it also includes feedback mechanisms at various stages.

However, actual implementations of scene interpretation algorithms do not consider all these principles. In particular, we have still a big, application dependent gap between the extracted features and the related object characteristics as described in the knowledge (or memory) representation. The necessary bridging process of *perceptual organisation* (like grouping of lines and areas into meaningful structures) has been neglected too much in the past and is mainly responsible for unsatisfying classification results (Schenk, 2003).

With the advent of multi-sensor systems as presented in the previous section, the number of extracted scene features increases. To deal with this large amount of information and eventually inherent contradictions, a *fusion* process on various levels becomes necessary. While most efforts in the past have been laid upon fusion on signal level, in our context a fusion on feature level is of great importance in order to integrate the several information in the sense of perceptual organisation as described above. Hence, in the following section we want to demonstrate a concept and an implementation of such a feature level fusion process for multi-sensor data for the particular application of extracting and classifying topographic surface edges.
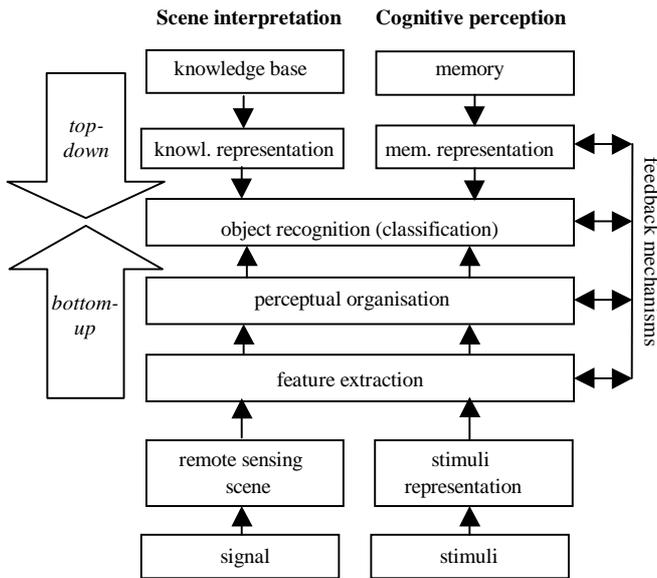

Figure 3.   Analogy between the processes of remote sensing scene interpretation and cognitive perception

4.   SURFACE EDGE EXTRACTION

4.1   Definitions

For the description and interpretation of topographical surfaces those curves are of major interest that represent either local maxima (ridge lines, watersheds), local minima (valley lines) or the border between surfaces with significantly different gradients. In the following we will concentrate on the latter type, which will be termed here *surface edges* and can be seen as a subset of the Digital Surface Model (DSM). Hence, surface edges combine the "hard" edges of topographical objects (*object edges*, like those of buildings or vegetation) and of *terrain edges* (as a subset of the Digital Terrain Model, DTM, like embankments, ditches, etc.). It shall be noted that the commonly used term of *breaklines* is strictly not correct because those represent only particular edges which had been generated through geomorphologic processes (Brunner, 1985).

Surface edges represent either an abrupt gradient change only, or they build complex objects which might consist of a lower and upper edge as well as a surface in between (like with an embankment). In principle surface edges are modelled by 3D-

vectors, whereas for a couple of applications (like topographical maps) a 2D-ground plan representation is sufficient. Unfortunately, from a modelling point of view, generally accepted quantitative criterions for surface edges (in particular thresholds for surface gradients) do not exist. As one example, the German ATKIS system specifies only the height and length of the object but no gradient value for capturing embankments.

4.2   Application potential

Surface edges can be seen as value added information and an improvement to any given Digital Elevation Model (DEM). Typical applications using edge information are for example flood prevention and river and drainage management, where characteristic lines for hydrological/hydraulic models are needed, the inspection survey of power lines, or the generation of 3D city models. Furthermore surface edges define the outline of so-called reduction surfaces which have to be masked out from DSMs in order to derive DTMs in the process of a DSM normalisation (Schiewe, 2003). Finally, they can significantly contribute to a reduction of data amount of very densely measured or interpolated DEMs.

4.3   Previous work

As regularly mentioned in the literature, the (semi-)automatic derivation of surface edges from irregularly or regularly spaced elevation points has led to unsatisfying results so far (e.g. Petzold et al., 1999; Pfeifer & Stadler, 2001). Obviously this is mainly due to the still limited quality of the input data in terms of the spatial resolution or point density as well as the geometrical accuracy in the vertical and horizontal components. On the other hand, these limitations will not be valid anymore in the near future with a certain probability, or can even be by-passed with some technical efforts nowadays. Hence, advancements with suitable algorithms for the extraction of surface edges are of great importance.

One of the major contributions for automatic surface edge detection in the past came from Wild et al. (1996) who applied an adaptive edge preserving filter in the process of the DEM generation before extracting edge pixels through gradient filters (e.g. a Sobel filter). Brügelmann (2000) used the second derivatives and hypothesis testing to derive regions of break points which then had to be further processed. Kraus & Pfeifer (2001) describe the derivation of 3D structure lines which uses the pre-knowledge of an approximate ground plan of the edges. However, it has be stated that these and other algorithms (like the ones of Chakreyavanich, 1991, or Gaisky, 2000), which are based upon geometrical information only, cannot compete with the manual, photogrammetrical measurement of surface edges in terms of completeness and accuracy.

4.4   Methodology

4.4.1   Core idea and outline: In contrast to other algorithms our proposed multi-sensor data fusion approach for the extraction of surface edges differs with respect to the following aspects:

- We will not only use one single elevation data set but also the various *multiple reflections* from a laser scanning systems (as presented in section 2.2) in order to increase

the amount of information that is introduced into the classification.

- We will not only use geometrical, i.e. elevation data, but also *semantical information* as derived from image data. This simulates the manual digitising process where thematic information about the underlying objects is introduced simultaneously and a categorisation of the extracted edges is performed.
- Instead of an edge-based approach we will apply a *region growing* algorithm, thus by-passing the problem of linking detected edge pixels to connected lines.

Figure 4 sketches the outline of the algorithm: A segmentation delivers borderlines which can be seen as candidates for surface edges. On base of the extracted features of the outline and the interior region the follow-up classification performs a hypothesis testing on these candidates as well as a categorisation into semantical classes (like walls or embankments). After the classified edges are converted into the vector domain, some post-processing steps (dilation by matching with image edges, and smoothing) are performed. At both stages, the classification and the post-processing, some grouping processes in the sense of perceptual organisation will be applied.
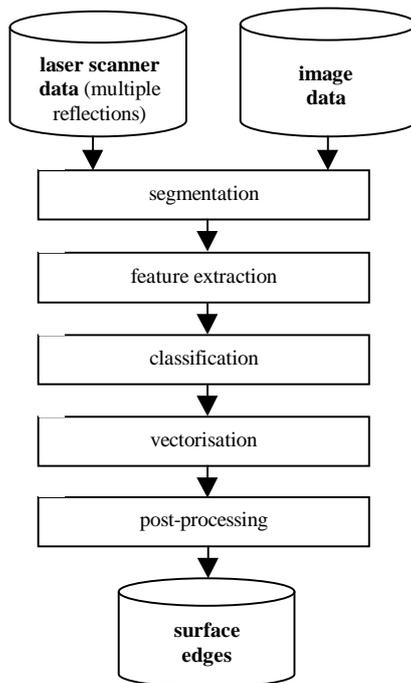

Figure 4. Outline of the proposed surface edge extraction approach using multi-sensor data

In the following sections the whole process will be explained in more detail by concentrating on the extraction of only one type of surface edges, namely building walls.

4.4.2   Segmentation: For the case under consideration, the detection of walls, both edge detection and region growing are aspects of the same processes under the assumption of step edges (Pavlidis & Liow, 1990). Hence, instead of an edge-based approach we will apply a *region growing* algorithm, thus by-passing the subsequent problem of linking candidate pixels, as detected by any edge filter, to connected lines. Furthermore,

the linkage between segments and attached borderlines is of great advantage for the subsequent classification process.

For the segmentation we use the software system *eCognition* (Baatz & Schäpe, 2000) which uses an extension of a region growing method called Fractal Net Evolution Approach (FNEA). As we want to extract only building walls at this stage of our study, we introduce the lowest elevation values from the last laser scanner echo (LE-low) as the heterogeneity feature for the segmentation. The LE-low represents only the ground surface and buildings (see section 2.2 and figure 2). Keep in mind that the LE-low reflection leads to an inside "buffering" of the real borderlines into the interior of the objects.

4.4.3   Classification of segments: The classification step performs not only a hypothesis testing of the surface edge candidates but also their categorisation into semantical classes. As already pointed out, in this study we want to concentrate on buildings walls only.

Our two-stage classification procedure starts with an *elimination approach*. Here features with rather "weak" threshold values are introduced which leads to a set of segments that classify nearly all buildings correctly (thus minimising the number of omission errors) but still include a considerable number of misclassifications (commission errors). For the classification of segments that are surrounded by walls we are considering the following feature values:

- Area greater than 50 m^2 – considering the minimum area of a building.
- Elevation difference to lower neighbours greater 3 m – considering the minimum height of a building.
- Normalised Difference Vegetation Index (NDVI) less than 0.05 – considering the relatively high reflections in the red spectrum and the relatively weak reflections in the near infrared spectrum due to roof colour and material.

Consequently, the goal of the second step has to be the reduction of the commission errors. In the following only the above obtained subset of candidates is taken into account. Because the object description using the following features is neither geometrically sharp nor standardised we introduce partial rather than crisp memberships, i.e. we apply a *fuzzy logic classification* approach. For the classification of segments that are surrounded by walls we are considering the following feature value ranges that shall make the distinction against vegetation segments and that are modelled by a linear membership function:

- NDVI (see above): Based on the hypothesis that with a smaller NDVI value the possibility of the existence of a building becomes larger, we introduce the value range between the minimum value (membership value $\mu=1$) and the above applied threshold of 0.05 ($\mu=0$).
- Rectangular fit: After creating a rectangle with the same area as the considered segment, the area of the object outside the rectangle is compared with the area inside the rectangle, which is not filled out with the object. For buildings a rather high value with a maximum of 1.0 can be expected. Thus the fuzzy value range extends from 0 ($\mu=0$) to 1 ($\mu=1$).
- Standard deviation of elevation: Due to some very high values at the edges (walls) we can expect high standard

deviations for buildings. Hence, the value range extends from the minimum variance ($\mu=0$) to the maximum ($\mu=1$).

Finally a combination of the single fuzzy values takes place through an unweighted averaging. Those segments, that have been classified as potential buildings in the first step and which also show here a combined membership value larger than 0.33, are now classified as buildings.

4.4.4 Post-Processing: At this point of the algorithm the **vectorisation** of the selected building outlines only is performed.

The problem with these *classified edges* is that they are based on the LE-low reflections of the laser scanning system which represent the outstanding objects in reduced size compared to the real outlines (see figure 2). Hence, a **dilation** has to be performed. In principle, multi-spectral imagery is able to locate these edges which more precisely (Schenk & Csatho, 2000). It has to be noted that these *images edges* would not have been a suitable input into the previous segmentation step because numerous edges of all objects would have been detected and several effects like gaps or over-sampling would have disturbed the object delineation. At this stage of our study we derive the image edges simply by using the above applied segmentation algorithm (figure 5, top left).

Because the image edges are rather imperfect, in the following the *edge matching* process is equivalent to a *buffering* of the classified edges into the direction of the outside image edges. In order to estimate the buffer distance we compute for every vertex of the classified edge the nearest distance to the surrounding image edge and build the average of those distances that are smaller than 1.5 pixels. This threshold is necessary in order to neglect incorrect or further image edges that are too far away from the real building outline. The value corresponds to the maximum positional error of the classified edges as derived from the LE-low reflections. The buffer operation yields, after dissolving barriers between the buffer boundaries, the *dilated classified edges* (figure 5, top right).
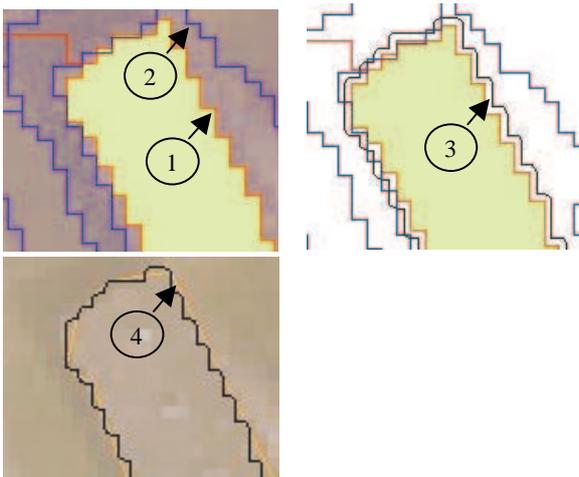

Figure 5. Post-Processing: top left: selected building (yellow) with classified edge (red, "1") and image edge (blue, "2"), top right: additionally dilated classified edge (black, "3"), bottom: smoothed dilated classified edge (orange, "4").

Finally a **smoothing** of these edges takes place using the Douglas-Peucker algorithm (figure 5, bottom).

4.5 First empirical results

With respect to the **segmentation** process the visual inspection yielded satisfying results with the selected grade of generalisation. The geometrical accuracy of the selected heterogeneity feature LE-low leads to a good separation of buildings from their surrounding objects.

Obviously, the key problem of the segmentation procedure is the proper and automatic choice of the grade of generalisation. It is possible to use multiple generalisation levels, for instance by applying the methodology of "Classification on multiple segment levels" (Schiewe, 2003). However, due to complexity reasons we have chosen only one scale level for this study.

The **classification** accuracy is determined through a visual interpretation of the obtained segments and expressed in terms of the error coefficient C as follows:

$$C = 1 - \frac{omission\_errors + commission\_errors}{total\_number\_of\_pixels\_in\_class}$$

Thus, the higher the coefficient C (with a maximum of +1.0) the better the overall classification accuracy is. After the first classification stage (elimination procedure, refer to section 4.4.3) C amounts to 0.88 where 95.3% of all buildings have been detected. The second stage (fuzzy logic classification) significantly reduces the number of commission errors so that the coefficient C becomes 0.92. While the remaining omission errors (4.7% of all buildings are not detected) are due to segmentation problems (see above), nearly all commission errors occur along the scene border. If additional knowledge would have been introduced here, the number of commission errors could have been reduced even more and the coefficient C would have been increased to 0.97.

It has to be stated that in principle the detection of buildings in FALCON data sets can be performed with an even increased accuracy if the imagery would have been taken during summer time. For our data set, which was captured in early March, the NDVI was not as effective as in other studies (e.g. Schiewe, 2003) for the distinction between buildings and vegetation (in particular deciduous trees).
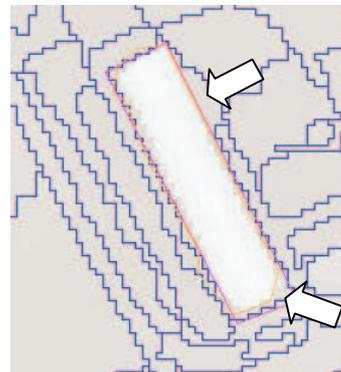

Figure 6. Result of post-processing step (orange line) compared to manually digitised edge (red rectangle) as well as numerous image edges (blue), overlaid onto LE-low elevation data.

Figure 6 gives a typical example of the effects of the **post-processing** operations. It demonstrates the necessity of both, the choice of elevation data for delineating buildings rather than image data (where shadows often extend or falsify the outline; indicated by the upper arrow), and the dilation operation which leads to a rather good coincidence with manually digitised edges. However, remaining errors occur which are mainly due to the input data quality (indicated by the lower arrow). It has to be noted that so far only few and simple post-processing operations have been applied.

5. CONCLUSIONS

Multi-sensor data yield the advantage of producing more scene features, however, the crucial point in the interpretation workflow is still the big, application dependent gap between these features and the related object characteristics. Hence, the **central goal** of this contribution was to present a methodology that bridges this gap for an important application, the extraction of topographic surface edges with the emphasis on building edges. In contrast to other methodologies for object edge extraction we propose an algorithm into which several features of both, the multiple reflections of the laser scanning system and the multi-spectral imagery, have been introduced. With that not only geometrical but also semantic information are used. As we concentrate on the derivation of building edges, we applied a *region growing* algorithm, which by-passes the problem of linking detected edge pixels to connected lines.

So far we have presented first qualitative **results** of our approach. It could be shown that the simultaneous usage of geometrical and semantic information definitively improves the classification and delineation of object edges compared to the single use of only one source. However, we have still to state some imperfect results which are mainly due to the quality of the input data in terms of the point density of the laser scanning elevations and the image acquisition date which has led to rather less predictive NDVI values.

In **future** several improvements of the methodological components will be addressed. For instance, we will apply a multi-scale segmentation instead of the use of a single segment layer. Additionally, the edge image matching process will be further developed by introducing even more features. Finally, the methodology will be extended in order to extract further surface edge classes (like embankments or ditches).

REFEERENCES

Baatz, M., & Schäpe, A., 2000. Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T. & Griesebner (Eds.): *Angewandte Geographische Informationsverarbeitung XII*, Wichmann, pp. 24-29.

Brügelmann, R., 2000. Automatic breakline detection from air-borne laser range data. In: *Int. Archives of Photogrammetry and Remote Sensing* (Amsterdam), 33(B3), pp. 109-116.

Brunner, K., 1986. Bruchkante oder Geländekante? *Bild-messung und Luftbildwesen,* 54(3): p. 132.

Chakreyavanich, U., 1991. Regular Grid DEM Data Compression by Using Zero-Crossings: The Automatic Breakline Detection method. PhD these (Ohio State University), Report OSU-DGSS, no. 412., 90 pp.

Gaisky, D., 2000. Geomorphologic improvement of DTMs especially as derived from laser scanner data. In: *Int. Archives of Photogrammetry and Remote Sensing* (Ljubljana), 32(6W8).

Kraus, K. & Pfeifer, N., 2001. Advanced DTM Generation from Lidar Data. In: *Int. Archives of Photogrammetry and Remote Sensing* (Annapolis, Maryland), 34(3/W4), pp. 23-30.

Löffler, G., 2003. Aspects of raster DEM data derived from laser measurements. In: *Int. Archives of Photogrammetry and Remote Sensing* (Dresden), 34(3/W13).

Lohr, U., 2003. Precise LIDAR DEM and True Ortho Photos. In: Fritsch, D. (Ed.): *Photogrammetric Week '03*, Wichmann, pp. 111-116.

Pavlidis, T. & Liow, Y.-T., 1990. Integrating Region Growing and Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3), pp. 225-233.

Petzold, B., Reiss. P. & Stössel, W., 1999. Laser scanning – surveying and mapping agencies are using a new technique for the derivation of digital terrain models. *ISPRS Journal of Photogrammetry & Remote Sensing*, 54, pp. 95-104.

Pfeifer, N. & Stadler, P., 2001. Derivation of digital terrain models in the SCOP++ environment. Proceedings OEEPE Workshop on Airborne Laserscanning and Interferometric SAR for Detailed Digital Elevation Models (Stockholm).

Schenk, T., 2003. From Data to Information and Knowledgde. In: Schiewe et al. (Eds.): Challenges in Geospatial Analysis, Integration and Visualization II. Proceedings of Joint ISPRS Workshop, Stuttgart (CD-ROM): 105-110.

Schenk, T. & Csatho, B., 2002. Fusion of LIDAR and aerial imagery for a more complete surface description. In: *Int. Archives of Photogrammetry and Remote Sensing* (Graz), 34(3A), pp. 310-317.

Schiewe, J., 2003. Integration of multi-sensor data for land-scape modeling using region-based approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57 (5-6), pp. 371-379.

Schiewe, J., 2004. Status and future perspectives of the application potential of digital airborne sensor systems. *International Journal of Applied Earth Observation and Geoinformation* (in Review).

Wild, D., Krzystek, P. & Madani, M., 1996. Automatic breakline detection using an edge preserving filter. In: *Int. Archives of Photogrammetry and Remote Sensing* (Wien), 31(B3), pp. 946-952.