Automatic Mapping of Settlement Areas Using a Knowledge-Based Image Interpretation System

Bernd-Michael Straub[1], Markus Gerke[1], and Martin Pahl[2]

[1] Institute of Photogrammetry and GeoInformation (IPI), University of Hannover, Nienburger Str. 1, D-30167 Hannover, Germany
{straub, gerke}@ipi.uni-hannover.de

[2] Institute of Communication Theory and Signal Processing (TNT), University of Hannover, Appelstr. 9A, D-30167 Hannover, Germany
geoaida@tnt.uni-hannover.de

Abstract. We introduce the knowledge-based image interpretation system GeoAIDA and give examples for an image operator, extracting trees from aerial imagery. Moreover we present a generic grouping approach, based on the Relative Neighborhood Graph. The application of the tree operator to a test site shows that the introduced approach for the delineation of trees using Active Contour Models leads to good results. The grouping algorithm is used in order to identify building rows. In the paper we shortly describe the theory of the image operator, and the performance of the whole system is demonstrated by means of examples. Results from a test area show that the information about building rows can be used for the enhancement of the building reconstruction.

1 Introduction

One main task when aerial images of settlement areas have to be interpreted automatically is the reconstruction of trees and buildings [7]. In this paper we present how object extraction from high-resolution aerial images can be supported by structural scene analysis. The knowledge-based image interpretation system GeoAIDA (**Geo** **A**utomatic **I**mage **D**ata **A**nalyser) [cf. 4] acts as a control unit: It calls *top-down* operators and the respective *bottom-up* operators. Finally it checks the consistency of the results. The *top-down* extraction modules "know" how the objects in the scene appear in the images and generate hypotheses of these objects. In the data-driven phase (*bottom-up*) the structural knowledge about the scene is used to refine these hypotheses.

After a short introduction of the concept of GeoAIDA in section 2 of the paper, we describe two types of operators in detail. In section 3.1 a *top-down* operator dealing with the extraction of single trees, an advancement of the approach introduced in [22]. Furthermore we introduce a *bottom-up* operator for the structural analysis of objects in section 3.2. It considers topological relations, represented by the Relative Neighborhood Graph [23], and geometric constraints in order to find rows of buildings. This information is then used to refine the reconstruction. Finally we present results to demonstrate the performance and potential of these two operators in section 4.

J.L. Crowley et al. (Eds.): ICVS 2003, LNCS 2626, pp. 355–364, 2003.

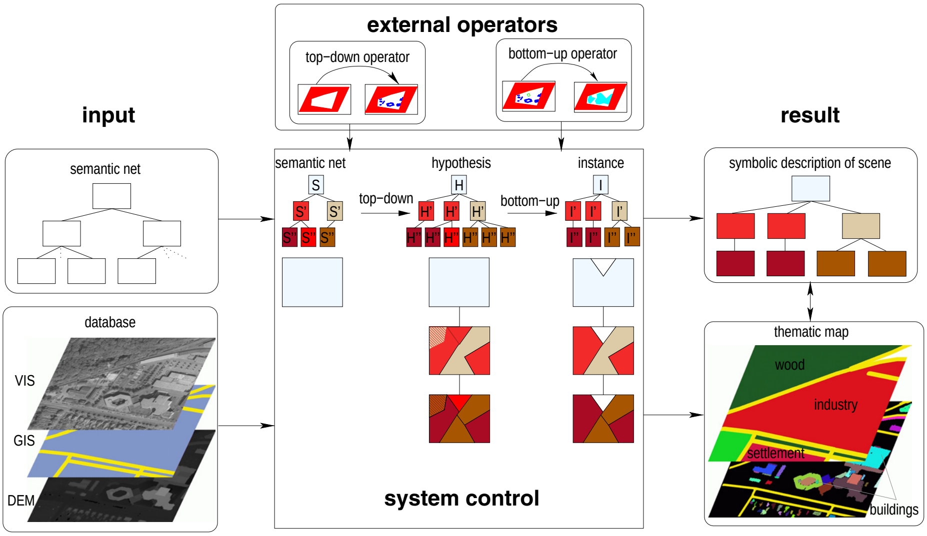

Fig. 1. GeoAIDA design

2 The Concept of GeoAIDA

Image interpretation systems are often restricted to segmentation and classification of images into one or few different classes. They are highly specialized and optimized for a certain task or have difficulties processing large images. Especially methods which follow a strict structural approach [14, 17, 24], i.e. working with primitive objects extracted from the image data, are not capable of handling whole aerial images due to the large number of extracted primitives in such images. GeoAIDA is designed for the automatic extraction of objects from remote sensing data. On the input side the system consists of the components *database* and *semantic net* (refer to Fig. 1 for an overview). Semantic nets provide a formalism for knowledge representation using nodes and links as basic elements. Nodes represent objects, while relations between these nodes are described by links. The a priori knowledge about the scene under investigation is stored in a generic semantic net. This net is called concept net and includes the objects expected in the input data. The nodes of the net are ordered strictly hierarchically, i.e. each node has exactly one parent node. Attributes can be assigned to each node dynamically, common attributes are *name*, *class* and the associated *top-down* and *bottom-up* operators.

The scene interpretation is handled by *top-down* and *bottom-up operators*, which are called by the system control unit. Results are shown in an interactive map which consists of a *symbolic description of the scene* and a *thematic map*.

The core *system control* queries the image database, reads the semantic net as well as project descriptions and generates hypotheses by calling *top-down* operators. A *top-down* operator associated with a node is capable of detecting objects of this kind of node (called *class*) in the given input data. For each object, which is expected in the scene, a hypothesis node is instantiated. The *bottom-up* operator investigates the relationship between the subnodes and groups them into consistent objects of the class given by the node the operator is connected to.

3 Interpretation Process

In this section two external operators for the extraction and grouping of objects are described. Three different semantic levels are used for the creation of the hypotheses: On the highest abstraction level the scene is segmented into *GroupOfTrees*, *BuildingAreas* and the background. The segmentation is based on thresholds in the Difference Normalized Vegetation Index (NDVI) and the normalized Digital Surface Model (DSM). The normalized DSM is the difference between the Terrain Model and the Surface Model, i.e. the ground topography is removed from the height dataset. Using these data sources as input information, the classification of the segmented regions leads to sufficient results in our test site. In more complex situations it can be useful to formulate the segmentation as a statistical classification problem; for example the use of a Markov random field is proposed by [2]. The detection of trees is performed together with the detection of buildings, in some cases as a pre-processing step for the building reconstruction, as described in [10]. An excellent overview of the recent research work in the domain of the automatic extraction of objects from remote sensing data is given in [3]. The second abstraction level is the object level. Based on the first segmentation the system calls a *top-down* operator for the instantiation of trees and buildings. Instances of *Trees* are generated as child nodes of the *GroupOfTrees* areas in the scene, and finally the contour of the tree's *Crown* – representing the third abstraction level – is delineated. This *top-down* operator is described in section 3.1 of this paper.

Instances of *Buildings* are created by the system according to the approach presented in [9] using a histogram analysis for the detection of buildings inside the *BuildingArea*, followed by the use of invariant geometric moments for the reconstruction of single buildings. The last step is the grouping of all the buildings in the scene, with the aim to find buildings arranged in a straight line. This is done with the *bottom-up* operator described in section 3.2.

Aerial color infrared (CIR) images and height data of a test site in Grangemouth, Scotland were used for the investigations. The CIR images were acquired in summer 2000 by the French company ISTAR. The image flight was carried out with 80% overlap along and across the flight direction using an analogue aerial film camera. The image scale is 1 : 5000, which leads to a GSD (Ground Sampling Distance) of 10 cm at a scanning resolution of 21 μm. Based on these images ISTAR has produced a DSM and a true orthoimage mainly automatically, using the approach described in [8]. The orthoimage and the DSM cover an area of 4 km^2, the GSD of the DSM is about 0.2 m.

3.1 A *top-down* Operator for the Extraction of Trees

The *top-down* operator for the extraction of trees is described in this section. The extraction is subdivided into two phases, first the detection of trees is performed, and in the second phase the crown perimeter is delineated. The precise delineation of the crown's contour is an important task within the reconstruction of a tree. The outline can be used for the estimation of the position and the radius of the stem; which are relevant parameters of a tree from a forestry point of view [11]. Furthermore, it is useful for a precise estimation of texture and color values without having the problem of mixed pixels, which is helpful for a more detailed classification of the type of the tree or its vitality.

The delineation of the individual crowns is done using Active Contour Models, also called Snakes. Snakes were introduced by Kass et al. [13] as mid-level algorithm which combines geometric and/or topologic constraints with the extraction of low-level features from images. The principal idea is to define a contour with the help of mechanic properties like elasticity and rigidity, to initialize this contour close to the boundary of the object one is looking for, and then let the contour move into the direction of the boundary of the objects. The original energy based approach for the contour can be reformulated to a pure geometry based approach, called Geodesic Active Contours [6]. Recent developments combine Geodesic Active Contours with level set methods [18], then the topology of the Active Contour can change during the optimization.

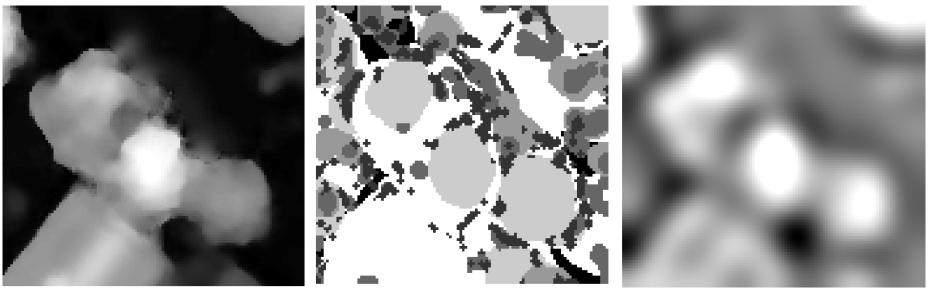

Fig. 2. The left image shows three different trees in the DSM. The optimal local scale is coded in the middle image, grey values assign the diameter of the corresponding region. The second derivative used for the extraction of the crown positions is printed on the right side using light grey for negative values and dark grey for positive values. The scale parameter used for the computation of the second derivative is selected according to the regions in the middle image.

In our approach one Active Contour Model is initialized for every tree as a circular shaped closed contour with an approximate value for the radius and the center position. The computation of these approximate values can be looked upon as the critical point in the whole extraction process.

A segmentation of the DSM into regions having the same diameter is used for the estimation of the radius. The segmentation is performed using morphological bandpass filters with a circular structuring function as proposed in [15]. The maximum value of the normalized response of the bandpass filters gives the local optimal scale for the appropriate pixel (Fig. 2). The optimal scale for the extraction of crowns in linear scale space [16] can be selected with this information about the diameter of the corresponding region. The sigma of the Gaussian - being the filter function in linear scale space - is proportional to the diameter of the region [20]. This property is used for the extraction of tree hypothesis. The minima of the DSM's second derivative in scale space are valid hypothesis of the searched tree positions, if the scale is selected correctly. These steps are depicted in Fig. 2: The left image shows the original DSM, the optimal local scale in morphological scale space is presented in the middle, and the respective second derivative in linear scale space is presented in the right image. The final result of the reconstruction using Active Contour Models is depicted in Fig. 4.

3.2 Grouping Approach Based on Neighborhood Graphs

In this section we present a generic approach based on neighborhood graphs for the grouping of objects and show how this approach can be used to create instances of building-rows. These rows are used to refine the orientation of the automatically extracted buildings.

Neighborhood graphs play an important role in many image analysis tasks and geographic applications [cf. 12, Ch. 7]. In the field of road extraction the network characteristic, i.e. the connection of single road segments to a road network represented by a graph is of substantial importance [cf. 21]. In [19] neighborhood relations between cars are used to detect parking lots. An other example is the clustering of geographic data to support generalization [1]. This clustering - provided by graph approaches - is necessary to obtain information about the structure of the data.

We use a Relative Neighborhood Graph (RNG) as introduced by Toussaint [23] to represent the topological relations between instances of the *Building* class; members of an arrangement like a row have to be connected by the edges of the RNG. We call constraints reflecting the topological relations of an arrangement *topological constraints*. In [23] the proximity definition of "relatively close" can be found:

> If we consider a set of distinct points $P = \{p_1, p_2, \dots, p_n\}$ in the plane two points p_i and p_j are supposed to be "relatively close" if $d(p_i, p_j) \leq \max[d(p_i, p_k), d(p_j, p_k)] \forall k = 1, \dots, n, k \neq i, j$, where d denotes the distance.

The edges of the RNG are connecting each pair of points which are "relatively close". Generally the RNG belongs to the family of *proximity graphs*; other representatives of this family are the Delaunay Triangulation (DT) or the Minimum Spanning Tree (MST). Whereas the DT is a superset of the RNG and the MST

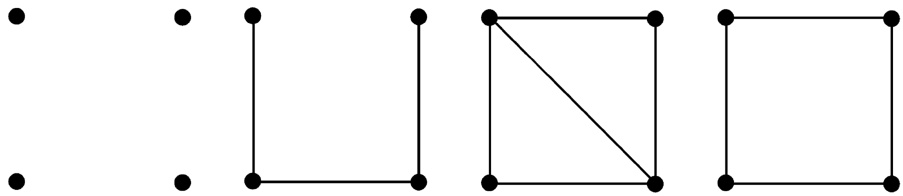

Fig. 3. Four points in the plane and the MST, DT and RNG of these points, from [23]

is a subset of it [23]. In Fig. 3 the differences between the three representatives of proximity graphs are pointed out: On the left side a set of four points in the plane is shown. The MST of these points is depicted in the second picture, besides the DT and the RNG. Our perception would "do" the same like the RNG and connect the four points in order to yield a rectangle. Besides the *topological constraints* we define geometric constraints which reflect the particular arrangement, e.g. for a row: "All instances have to be aligned to a straight line". We call these constraints *global geometric constraints.* Additionally some *local geometric constraints* may be formulated, for example "all buildings in a row must have the same orientation".

We use the RNG as an adequate representative of the topology. The distance between two buildings is the shortest distance between the appropriate region. This distance is assigned to the edges of the RNG as a weight. In order to express the condition that buildings may not exceed a given distance from each other a maximum weight for the edges can be required. Additionally the number n_s of nodes in this graph can be restricted, and a minimum can be required. We define the minimum number of buildings in a row is $n_{s_{\min}} = 3$. Then we demand that the neighbored buildings should lie on a straight line. This is a global geometric constraint, because it concerns all buildings in one row. Every buildings center of gravity may not exceed a maximum distance $d_{\max}$ to a common straight line. This maximum distance $d_{\max}$ is set to $d_{\max} = 3$m. Additionally we want all buildings belonging to a row of buildings to have the same orientation, this is a local geometric constraint.

4 Results

In this chapter we show some results of the introduced operators. The test site on which we concentrate is a small part of the Grangemouth data set (cf. Sec. 3).

4.1 Results of Tree Extraction

The results of the described algorithm are depicted in Fig. 4. On the left image the perimeter of a *GroupOfTree* instance is superimposed to the optical image.

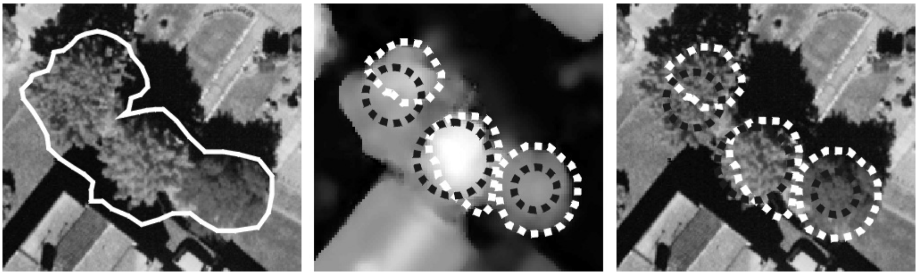

Fig. 4. Exemplary results of the tree extraction operator are depicted in these three pictures

Fig. 5. Reconstructed buildings with marked centers of gravity and row of buildings (left), buildings after orientation enhancement (right)

The initial contour for the Snakes are depicted as black dotted lines, and the final solution as white dotted lines superimposed onto the DSM in the center of Fig. 4 and onto the optical image to the right.

Three typical cases can be explained with the help of this small group of three trees. The left crown in Fig. 4 consists of a clear "mountain", which is correctly delineated in the DSM, but the optical image shows, that the crown is larger than this one "mountain". The middle crown is not extracted correctly, the lower left part of the contour is too far away from the center. The right tree in the example group is correct. Another case, which is not depicted here, occurs sometimes: If the initialization of the Snake is bad, the contours contracts to a length of zero.

4.2 Results of Building Extraction and Building Row Identification

We applied our approach for building extraction to the used image data. For more detailed results concerning the building extraction operator see [9]. The results for our subset are depicted on the left image in Fig. 5. Besides the outlines of reconstructed buildings, the centers of gravity of the buildings and the identified row of buildings is displayed. The orientation of buildings is needed for the fulfillment of a local geometric constraints. Now, after identifying the rows, it can be enhanced. This can be observed in the images: In the original reconstruction result the orientations of the buildings in the row oscillate. After the identification of the row all concerned buildings are oriented accordingly to a common average. This leads to the outlines depicted in the right image in Fig. 5.

The identification of rows matches the reality very well, the orientation manipulation results in a significant improvement of the reconstruction result.

5 Conclusions

We have presented an approach for the automatic extraction of trees in settlement areas and a generic grouping algorithm using neighborhood relations. The GeoAIDA-System allows the combination of several image analysis operators together with structural analysis. The process of the generation of hypothesis and their evaluation can be clearly separated. Moreover the possibility to integrate external operators makes the system modular and allows the combination of very complex operators. In the near future GeoAIDA will be utilized in the framework of a project initiated by the German Federal Agency for Cartography and Geodesy (BKG). In this project a system for the image based automated quality control of the Germany-wide available topographic vector data set ATKIS DLM-Basis is being developed [5]. Here GeoAIDA will be the main system integrating several operators.

While the essential part of the algorithm for the extraction of buildings was introduced in [9] the algorithm for tree extraction as described in this paper is a further development of the algorithm presented in [22]. The example for the grouping approach shows that the building reconstruction process can be supported by the structural analysis. The process of building orientation adjustment after identifying a building row obviously leads to an improved reconstruction result. Nevertheless a verification of building edges in the image data is indispensable for a further result enhancement. In the future we will test our grouping approach using a larger dataset in order to learn more about the algorithms limitations. The Active Contour Model for the delineation gives the possibility to introduce knowledge about the object *Tree* on a very general level. The problem of finding approximate values for the initialization of the Snake was solved by means of the estimation of the local optimal scale in morphological scale space. The interrelationship between the optimal scale in morphological and in linear scale space was used for an estimation of the optimal position and radius of tree-crowns.

Acknowledgment. Parts of this work were developed within the IST Program CROSSES financed by the European Commission under the project number

IST-1999-10510. The development of GeoAIDA is partly funded by BKG. Aerial Images, DEM and True Orthoimages © ISTAR. DEM and True Orthoimages made by ISTAR.

References

[1] Anders, K.H., Sester, M.: Parameter-Free Cluster Detection in Spatial Databases and its Application to Typification. In: International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII (A4), Amsterdam (2000) 75–82

[2] Baillard, C., Maitre, H.: 3-D Reconstruction of Urban Scenes from Aerial Stereo Imagery: A Focusing Strategy. In: Computer Vision and Image Understanding. **76**, No. 3. (1999) 244–258

[3] Baltsavias, E., Grün, A., van Gool, L.: Automatic Extraction of Man-Made Objects from Aerial and Space Images (III). Balkema Publishers, Rotterdam (2001) 415 pages

[4] Bückner, J., Pahl, M., Stahlhut, O., Liedtke, C.-E.: A Knowledge-Based System for Context Dependent Evaluation of Remote Sensing Data. In: Pattern Recognition, 24th DAGM Symposium, Nr. 2449 in LNCS. Springer, Zurich, Switzerland (2002) 58–67

[5] Busch, A., Willrich, F.: Quality Management of ATKIS Data. In: Proceedings of OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management. Istanbul (2002) 30–42

[6] Casselles, V., Kimmel, R., Sapiro, G.: Geodesic Active Contours. International Journal of Computer Vision. **22**. (1997) 61–79

[7] Fuchs, C., Gülch, E., Förstner, W.: OEEPE Survey on 3D-City Models. In: OEEPE Publications, Bundesamt für Kartographie und Geodäsie, No. 25, Frankfurt (1998) 9–123

[8] Gabet, L., Giraudon, G., Renouard, L.: Construction automatique de modèles numériques de terrain haute résolution en milieu urbain. Société Française de Photogrammétrie et Télédétection. **135**. (1994) 9–25

[9] Gerke M., Heipke C., Straub B.-M.: Building Extraction From Aerial Imagery Using a Generic Scene Model and Invariant Geometric Moments. In: Proceedings of the IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, IEEE Piscataway (2001) 85–89

[10] Haala, N., Brenner, C.: Extraction of Buildings and Trees in Urban Environments. In: ISPRS Journal of Photogrammetry and Remote Sensing. **54**, No. 2-3. (1999) 130–137

[11] Hyyppä, J., Hyyppä, H., Ruppert, G.: Automatic Derivation of Features to Forest Stand Attributes Using Laser Scanner Data. In: International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII (B3), Amsterdam (2000) 421–428

[12] Jaromczyk, J., Toussaint, G.: Relative neighborhood graphs and their relatives. In: Proceedings IEEE. **80**, No. 9. (1992) 1502–1517

[13] Kass, M., Witkin, A., Terzopoulus, D.: Snakes: Active Contour Models. In: International Journal of Computer Vision. **32**, No. 1. (1988) 321–331

[14] Kummert, F., Niemann, H., Prechtel, R., Sagerer, G.: Control and explanation in a signal understanding environment. In: Signal Processing, special issue on 'Intelligent Systems for Signal and Image Understanding'. **32**. (1993) 111–145

[15] Köthe, U.: Local Appropriate Scale in Morphological Scale-Space. In: R. C. B. Buxton(ed.): 4th European Conference on Computer Vision (1). Springer, Lecture Notes in Computer Science 1064. Berlin (1996) 219–228

[16] Lindeberg, T.: Scale-Space Theory in Computer Vision. Kluwer Academic Publishers, Boston (1994) 423 pages
[17] Niemann, H., Sagerer, G., Schröder, S., Kummert, F.: ERNEST: A Semantic Network System for Pattern Understanding. In: IEEE Trans. on Pattern Analysis and Machine Intelligence. **12**, No. 9. (1990) 883–905
[18] Paragios, N., Deriche, R.: Geodesic Active Regions and Level Set Methods for Supervised Texture Segmentation. In: International Journal of Computer Vision. **46**, No. 3. (2002) 223–247
[19] Quint, F.: MOSES: A Structural Approach to Aerial Image Understanding. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhäuser, Basel. (1997) 323–332
[20] Steger, C.: Extraction of Curved Lines from Images, In: 13th International Conference on Pattern Recognition. (1996) 251–255
[21] Steger, C., Mayer, H., Radig, B.: The Role of Grouping for Road Extraction. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images (II). Birkhäuser, Basel. (1997) 245–256
[22] Straub, B.-M., Heipke, C.: Automatic Extraction of Trees for 3D-City Models from Images and Height Data. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images (III). Balkema Publishers, Rotterdam. (2001) 267–277
[23] Toussaint, G.: The relative neighbourhood graph of a finite planar set. Pattern Recognition. **12**. (1980) 261–268
[24] Tönjes, R., Growe, S., Bückner, J., Liedtke, C.-E.: Knowledge-Based Interpretation of Remote Sensing Images Using Semantic Nets. Photogrammetric Engineering and Remote Sensing. **65**, No. 7. (1999) 811–821