# SCENE ANALYSIS IN URBAN AREAS USING A KNOWLEDGE-BASED INTERPRETATION SYSTEM

**Markus Gerke**
Institute for Photogrammetry and GeoInformation, University of Hannover
Nienburger Str. 1, 30167 Hannover, Germany - gerke@ipi.uni-hannover.de

Working Group III/4

**KEY WORDS:** Analysis, Automation, Building, Detection, Hierarchical, Integration, Urban

**ABSTRACT**

In this contribution the way algorithms for object detection in urban areas are integrated into the knowledge-based image interpretation system GeoAIDA is described. Generic scene models are used for object detection in settlement areas, whereas the implementation of the respective algorithms is a collection of stand-alone-operators.
With GeoAIDA a system is available which uses these operators in the first phase (model-driven, *top-down*) in order to generate hypotheses for objects in the scene. In the second phase (data-driven, *bottom-up*) the hypotheses are further processed using structural knowledge about the scene. Here, detected buildings are grouped using the Relative Neighborhood Graph. An example shows that the combination of low-level image operators and high-level grouping operators leads to enhanced scene analysis results.

## 1 INTRODUCTION

Our recent publications presented some approaches on automatic extraction of buildings and trees from aerial imagery. This work was embedded in the CROSSES project (CRO, 2002) and led to convincing results in this field of image interpretation.
In this paper the integration of image analysis algorithms into the knowledge-based image interpretation system GeoAIDA (**Geo** **A**utomatic **I**mage **D**ata **A**nalyser) is demonstrated. This software is being developed by the Institute of Communication Theory and Signal Processing (TNT), University of Hannover. For a detailed system overview see (Bückner et al., 2002). The system contains a knowledge-base, represented by means of a semantic network. Image operators are attached to the nodes of the network in order to find evidence of the objects in the scene. The image processing is done in the model-driven phase (*top-down*) by *top-down* operators. The objects, resulting from this process are called *hypotheses*. Moreover, the system allows the incorporation of structural knowledge about the scene, i.e. knowledge on how a multitude of objects can be grouped to parent objects. This process is data-driven (*bottom-up*) and done by so called *bottom-up* operators. Possible ambiguities, occurring in the *top-down* process are also solved in the *bottom-up* phase. The final objects are called *instances*.
The integration of such operators in a system like GeoAIDA is of vital importance because of the following reasons; firstly, the use of structural knowledge about the scene, taking into account several object classes is simplified when using an integrated system. Secondly, the development and study of new operators is supported by this system as low-level image operators can be combined with high-level operators, making use of structural knowledge. Last but not least the acceptance of automatic image analysis approaches increases if potential end users have a user-friendly interface.

The building detection operator as introduced in (Gerke et al., 2001) was exemplary integrated in GeoAIDA. Besides the detection of *Buildings*, an operator for the detection of *GroupOfTrees* was implemented, cf. (Straub and Heipke, 2001). Furthermore, a *bottom-up* operator for the identification of *Building-Groups* was applied.
Results for a test area show that the integration of image operators and the use of structural knowledge lead to an enhanced scene analysis.

## 2 AN EXEMPLARY GENERIC SCENE MODEL

Below a scene model, represented by a semantic network is formulated (figure 1). The topmost node of the network is called *Scene*. This node initializes the *top-down* process and it is the last node processed in the *bottom-up* step. The next level contains the concept *Settlement*, whereas this node contains the concepts *GroupOfTrees* and *Building-Group*. A *top-down* operator is assigned to the first concept. This means, an external program is called by this node in order to search for evidence of trees in image data. This operator is described in chapter 3.1. The *Building-Group* node contains one or more *Buildings*. Such grouping can be used later on as a preprocessing step for a detailed structural analysis, e.g. finding geometric arrangements of buildings. The grouping is carried out using the distance a *Building* has to its neighbor. The *bottom-up* operator attached to the *Building-Group* node is explained in chapter 4 as well as the *top-down* operator for *Building* detection in section 3.2.

## 3 IMAGE OPERATORS

In this chapter the image operators for the detection of *GroupOfTrees* and *Buildings* in aerial images are described. The term “detection” means in this context the assignment of image parts belonging to an instance of the particular

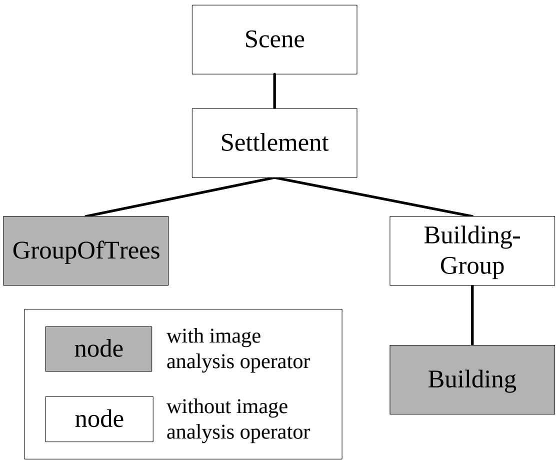


Figure 1: Exemplary Semantic Network

object-class, according to the underlying model. This analysis results in a label image.
For the description of the operators a 3-layer scene model is used (refer to figures 2 and 3). The topmost layer is called *Real World*. This layer contains the objects one wants to describe. Below the *Real World* the *Material and Geometry* layer is introduced. This one contains the physical properties of the objects and is data independent. The bottom-layer is called *Image*. The term "image" includes all possible raster data, e.g. optical images or surface models. During its *Life Cycle* an object calls algorithms which extract low-level features from images (*Create-Feature*) and it *creates instances* of other objects using these features. In the next two subsections the *Create-Feature* and *Create-Instance* operations for the objects *GroupOfTrees* and *Building* are explained.

### 3.1 Detection of GroupOfTrees

The detection of *GroupOfTrees* objects is comparable to the pixel-based classification as shown in (Haala and Brenner, 1999). For the differentiation between vegetated and non-vegetated (sealed) regions the radiometric information from the CIR-Image, namely the *NDVI*, is used. Then *Vegetation* regions can be found (① in fig. 2). The threshold value for the segmentation into vegetation and non-vegetation is determined using a histogram analysis. The discrimination between objects on and above the ground (*3D Region*) is done using a threshold in the normalized surface model (② in fig. 2). The instances of *GroupOfTrees* are created by means of an intersection between *Vegetation* and *3D Region* (③ in figure 2). For more details and a description for single tree extraction refer to (Straub and Heipke, 2001).

### 3.2 Detection of Buildings

In (Gerke et al., 2001) an approach for single building extraction using a surface model and a true CIR-orthoimage is described. In comparison to previous presentations the complexity of the problem is reduced as the reconstruction step is not contained: here image segments, containing single buildings, are of interest and not the vectorial description of the buildings.

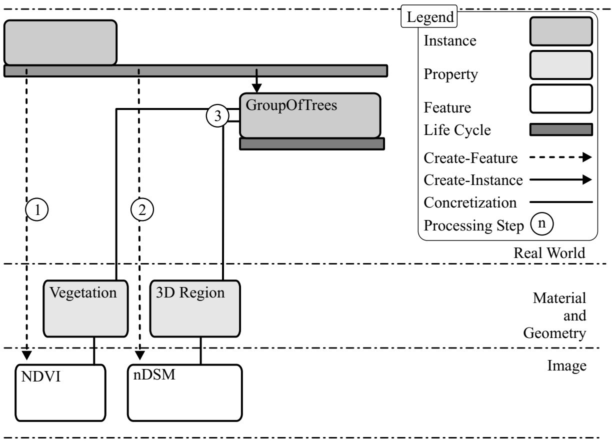


Figure 2: *GroupOfTree*-Detection Operator

The design of the implemented operator is depicted in figure 3. Similar to the detection of *GroupOfTrees* the analysis starts with the low-level analysis in the *NDVI* and the normalized surface model leading to *Non-Vegetation* and *3DRegions*. The intersection of these regions leads to *BuildingArea*-objects, refer to ①, ② and ③ in fig. 3. A *BuildingArea* contains one or more *Buildings* and *Cast-*

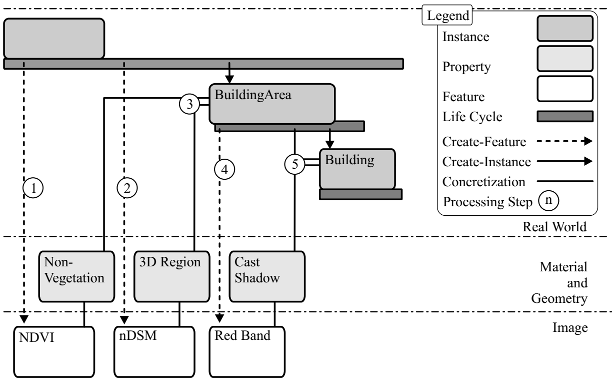


Figure 3: *Building*-Detection Operator

*Shadow* areas. This is because the *3DRegions* are often enlarged in the direction of cast shadows, formed by buildings. This can be explained by the generation process of the surface model: This was derived by matching algorithms, which can lead to poor results in cast shadow regions, cf. (Haala, 1996). The enlargement of *3DRegions* leads to a fusion of buildings in the surface model if they are connected by shadow. In order to separate the single *Buildings* from the *CastShadow* each *BuildingArea*-object carries out a histogram analysis in its domain of the red-band and finally creates instances of one or more *Building*-objects (④ and ⑤ in fig. 3), refer to (Gerke et al., 2001).

## 4 GROUPING APPROACH

After GeoAIDA called the *top-down* operators the image analysis is completed and the *bottom-up* phase, i.e. the data-driven-phase begins. Here the structural knowledge about the scene can be used for solving ambiguities, emanating from the *top-down* phase, based on an evaluation of

the image analysis results. This evaluation can be used for the propagation of higher-level object classes. An example for the decision whether an accumulation of different objects belongs to a settlement area or an industrial site is given in (Bückner et al., 2002).
In this example the focus is on the detection of *Building-Groups*. This means, all *Buildings*, detected by the image operator introduced above, are further analyzed in the sense that neighboring buildings are grouped together to an instance of the named concept. This grouping can be used to support the detection of geometric arrangements of buildings, such as rows.
For this approach the Relative Neighborhood Graph (RNG) as introduced in (Toussaint, 1980) is applied:

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

The edges of the RNG are connecting each pair of points which are "relatively close". A graph is weighted if real numbers are assigned to the edges. In this application a RNG of *Building* objects is to be formulated. The pointwise representation of regions (e.g. their centers of gravity) is not a very good choice in this case as the decision whether two *Buildings* are "relatively close" or not is better established in relation to the distances between the regions' borders. Therefore the distance between two regions is to be the shortest distance between the contours of these two regions, which can be calculated e.g. using the Hausdorff distance, cf. (Soille, 1999, Ch. 3.11.3).
If this distance is assigned to the edges of the graph, the *Building-Group* can be defined:

> Members of a *Building-Group* belong to a subgraph of the RNG of all *Buildings*, whereas the weight, assigned to each edge of this subgraph does not exceed a given maximum. The minimum number of *Buildings* belonging to a *Building-Group* is $1$. Each *Building* belongs to exactly one *Building-Group*.

This definition is implemented in the *bottom-up* operator. It gets the single *Building* regions from GeoAIDA and after the analysis it returns the description of the *Building-Group* instances by means of the assigned *Buildings* IDs.

## 5 RESULTS

The described semantic network with the *top-down* and *bottom-up* operators assigned to the respective nodes was entered in GeoAIDA. For the investigation, image and height data of a test area in Grangemouth, Scotland are used. The color infrared aerial images were acquired in summer 2000 for the CROSSES project (CRO, 2002). The image flight was carried out with 80% overlap along and across the flight direction. The image scale is 1:5000, which leads to a GSD of 10 cm at a scanning resolution of 21 $\mu$m. Based on these images a DSM and a true orthoimage were automatically derived by the French company ISTAR (Gabet et al., 1994). The orthoimage and the DSM cover an area of 4 km$^2$. A large part of the whole test site belongs to an industrial plant with sparse vegetation. A subset of the data with typical suburban characteristics was selected, as depicted in figure 4. The minimum height for a *Building*


Figure 4: CIR True Orthoimage Showing the Test Area

was set to 4 m. The maximum distance two *Buildings* may have in order to be members of the same *Building-Group* is 10 m. This is a heuristic value, but one can argue that such groups of buildings in an urban environment are separated by a street, which normally has a minimum width of approx. 10 m. The left part of fig. 5 shows a snapshot of the *Scene Viewer*. In this browser the *instances* are listed. All attributes of the objects like e.g. class-name, coordinates of the bounding box in object-space or user-defined attributes (given by *top-down* or *bottom-up* operators) are shown in the right column of this viewer. The *Scene Viewer* is linked to the *Result Map* as shown in the right image. In this map the label-image of the *instances* is shown, *Buildings* assigned to the same *Building-Group* are linked. The result of the object detection is as follows: The *GroupOfTree*-instances cover around 96% of the total tree area, but also approx. 12m$^2$ of the area where no trees are situated, and 100% of the 58 buildings were detected, but the two buildings in the northern part belong to the same *Building* instance. The result of *Building-Group*-detection is as expected as it corresponds to the formulated model.
The result of the analysis, represented in an *instance-net* is stored in a common *XML-file*. Therefore these results can be used by other programs, for example for drawing comparisons with reference data.

## 6 CONCLUSIONS

This paper demonstrates how image analysis operators can be efficiently integrated in a knowledge-based image in-


Figure 5: GeoAIDA-Scene Viewer and Result Map

terpretation system. It was shown how detected *Buildings* were successfully grouped to instances of *Building-Groups*. Our further work will increasingly focus on the structural analysis. The consideration of geometric conditions can lead to the detection of geometric arrangements of objects, such as rows or circles. The advantage of this procedure is twofold: On the one hand the detection of such arrangements supports the detection of other objects. For example one can argue that in urban environments streets are in general parallel to building-rows. On the other hand buildings, belonging to a building-row often are similar (e.g. they have the same orientation). This observation can be used to enhance the building reconstruction. The integration of operators into a system like GeoAIDA does support this further work as the solving of ambiguities, caused by competing hypotheses is done by the system.

## ACKNOWLEDGMENT

This work was partly developed within the IST Project CROSSES financed by the European Commission under the project number IST-1999-10510. The development of GeoAIDA is partly funded by the German Federal Agency for Cartography and Geodesy (BKG). Parts of this contribution were submitted for a special issue of the ISPRS Journal of Photogrammetry and Remote Sensing, planned to appear in April 2003.

## REFERENCES

Bückner, J., Müller, S., Pahl, M. and Stahlhut, O., 2002. Semantic interpretation of remote sensing data. In: Proceedings of Photogrammetric Computer Vision: ISPRS, Com III. Symposium 2002, Graz, *to appear*.

CRO, 2002. The CROSSES-Website. http://crosses.matrasi-tls.fr/ *(May - 01 - 2002)*.

Gabet, L., Giraudon, G. and Renouard, L., 1994. Construction automatique de modèles numériques de terrain haute résolution en milieu urbain. Société Française de Photogrammétrie et Télédétection 135, pp. 9–25.

Gerke, M., Heipke, C. and Straub, B.-M., 2001. 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, Rome.

Haala, N., 1996. Gebäuderekonstruktion durch Kombination von Bild- und Höhendaten. PhD thesis, Deutsche Geodätische Kommission. Series C. Vol. 460.

Haala, N. and Brenner, C., 1999. Extraction of buildings and trees in urban environments. ISPRS Journal of Photogrammetry and Remote Sensing 54(2-3), pp. 130–137.

Soille, P., 1999. Morphological Image Analysis. Springer Verlag.

Straub, B.-M. and Heipke, C., 2001. Automatic extraction of trees for 3d-city models from images and height data. In: E. Baltsavias, A. Gruen and L. van Gool (eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images, Vol. III, Balkema Publishers, Rotterdam, pp. 267–277.

Toussaint, G., 1980. The relative neighbourhood graph of a finite planar set. Pattern Recognition 12, pp. 261–268.