Automatic Detection of Buildings and Trees from Aerial Imagery Using Different Levels of Abstraction

MARKUS GERKE, BERND-MICHAEL STRAUB & ANDREAS KOCH[1]

Zusammenfassung: *In diesem Beitrag wird die Extraktion von Bäumen und Gebäuden aus hochauflösenden Luftbildern und Höhendaten im städtischen Bereich beschrieben. Der vorgestellte Ansatz nutzt zur Steuerung der Szeneninterpretation verschiedene Abstraktionsebenen, ähnlich der in GIS Objektmodellen verwendeten hierarchischen Strukturierung von topographischen Objekten.*

Abstract: *In this paper we describe our work on the automatic extraction of buildings and trees from aerial images and height data in urban environments. The control strategy of the presented approach is based on different abstraction levels. The abstraction levels are similar to the hierarchical structure of topographic objects in GIS object models.*

1 Introduction

In this paper we describe our current work on automatic detection and extraction of trees and buildings from aerial CIR orthoimages and normalized digital surface models. The main focus of our work is on the extraction of buildings and trees in an urban environment. For our work the application consists in building a simulation system for training emergency force officers to better respond to crisis situations, refer (CROSSES, 2001). One aim of our work is the automatic production of a highly detailed city model, which can be used in a wide variety of applications like games, tourist applications, and simulation computations. These kind of applications demand inexpensive data, and it is often possible to decrease the data quality requirements. These aspects argue for the application of automatic procedures for the object extraction. The paper is structured as follows: After a short literature review, we present the employed strategy of our approach. Afterwards we describe how this strategy is applied to the extraction of trees and building ground plans in an urban environment including examples of the obtained results. The paper concludes with an outlook onto future developments.

2 Related work

In this paragraph a brief literature overview concerning building reconstruction from image data is given. Currently we focus on reconstructing building shapes in the 2D plane. Therefore the term building shape is used in this paper in order to distinguish it from 3D building reconstruction. In consequence of using aerial data the building shape is approximated by the roof shape. WEIDNER (1997) introduces two approaches for building shape reconstruction. A segmented region is either adapted to a set of given prismatic building hypothesizes (which are represented by polygons in 2D) using the principle of the minimum description length (mdl, cf. BRUNN ET AL., 1995) or described by a parametric building model which is a rectangle in the 2D-plane. The advantage of fitting a polygon to the region is that it is the more comprehensive approach but on the other side the quality of the result depends on a sensitive tuning of control parameters. MAAS (1999) applied the analysis of invariant geometric moments in order to achieve simple parametric building models, which is similar to the approach presented by WEIDNER (1997). One advantage of using moments is that they lead to the five parameters (width, length, orientation and position) fixing the rectangle. The extraction of individual trees is also addressed by a number of authors. BRUNN & WEIDNER (1997) pro-

[1] Institut für Photogrammetrie und GeoInformation, Universität Hannover, Nienburger Straße 1, 30167 Hannover, e-mail: {gerke,straub,koch}@ipi.uni-hannover.de

posed to use the variance of surface normals to detect vegetation regions in height data. Laser scanner data and a color infrared image are used in combination by BRENNER & HAALA (1999) for the detection of vegetation areas in an urban environment by means of an unsupervised classification algorithm. POLLOCK (1994) proposed a generic model for a tree divided into a geometric part, a function for the leaf density distribution, and an illumination model. BRANDTBERG & WALTER (1998) have developed an approach for the extraction of trees from aerial images based on the curvature of edges in multiple scales.
Similar to BRENNER & HAALA (1999) we combine CIR and height data. For the delineation of individual trees we apply a morphological approach, which makes use of the circular shape of a tree crown. In the case of building shape reconstruction we decided to extend the approach presented in (MAAS, 1999), whereas we make use of its property to directly achieve the parameters of a rectangle. The restriction regarding the complexity of the building shape is not critical as our approach is applied to a suburban environment, where mainly simple shaped buildings can be found.

3 Hierarchical Image Analysis

The existing methods of image analysis can be mapped into a two dimensional solution space, following a proposition made by SUETENS ET AL. (1992). One axis describes the suitability for complex models and the other axis the suitability for complex image data. Methods for processing either simple models or simple images are available. However, complex models in conjunction with complex image data cannot be handled appropriately with existing methods (HEIPKE ET AL., 2000). Thus, a reduction in complexity is necessary. Since the model complexity is usually dictated by the application, only the image complexity can be manipulated. It can be reduced by means of a scale-space transformation (LINDEBERG, 1994) and by a reduction of the processing domain. We use a hierarchical model, which is structured by abstraction levels of the objects, as a kind of control strategy for our approach. An example for an object hierarchy is given in figure 1. The highest level of abstraction is given by the landscape, level 2 contains the super classes, different areas are placed in level 3, the next level consists of the objects and the lowest level of abstraction is given by the level of object components.

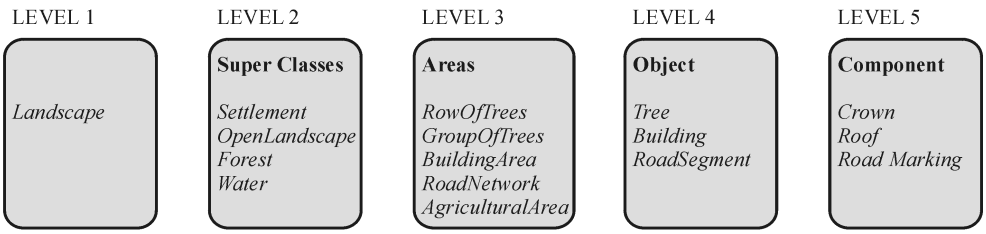

Figure 1: Different abstraction levels in the hierarchical structure of the landscape

In order to solve a problem with complex models and complex images, it can be argued that the following hierarchical strategy is feasible. At some level L of the hierarchy the image complexity is reduced using a scale-space transformation. The related object classes of level L+1 are searched for only in the parts of the scene, in which the context –given by the instances of the level L objects - is known. This procedure is repeated until the last level is reached, leading to a stepwise reduction of the complexity of the images. Therefore, simple object models can be applied in the actual level to be processed. The context always defines the domain of the scene relevant for the actual level and is also used to set initial parameters for the object extraction. The term s*cene* is used for the domain of the real world which is mapped in the images, the *scene description* consists of the instances of objects visible in the scene.

4 Tree and Building Extraction

The strategy is applied as described in this sequence. The context from level 2 –settlement- is known by means of the application, in general this context can be automatically extracted from coarse scale imagery (STRAUB ET AL., 2000). In the given settlement domain of the scene, areas –level 3- are extracted (see section 3.1), and inside the areas we focus on trees or buildings depending on the semantic of the given area. An overview of all processing steps is given in Figure 2, in this sequence diagram the different objects are ordered by means of their life cycle - from left to the right - and by the abstraction level from top to bottom. The "Material and Geometry" layer is used for class properties, which are independent of the abstraction levels. The “Image” layer contains the low-level features, which are extracted from the image data.

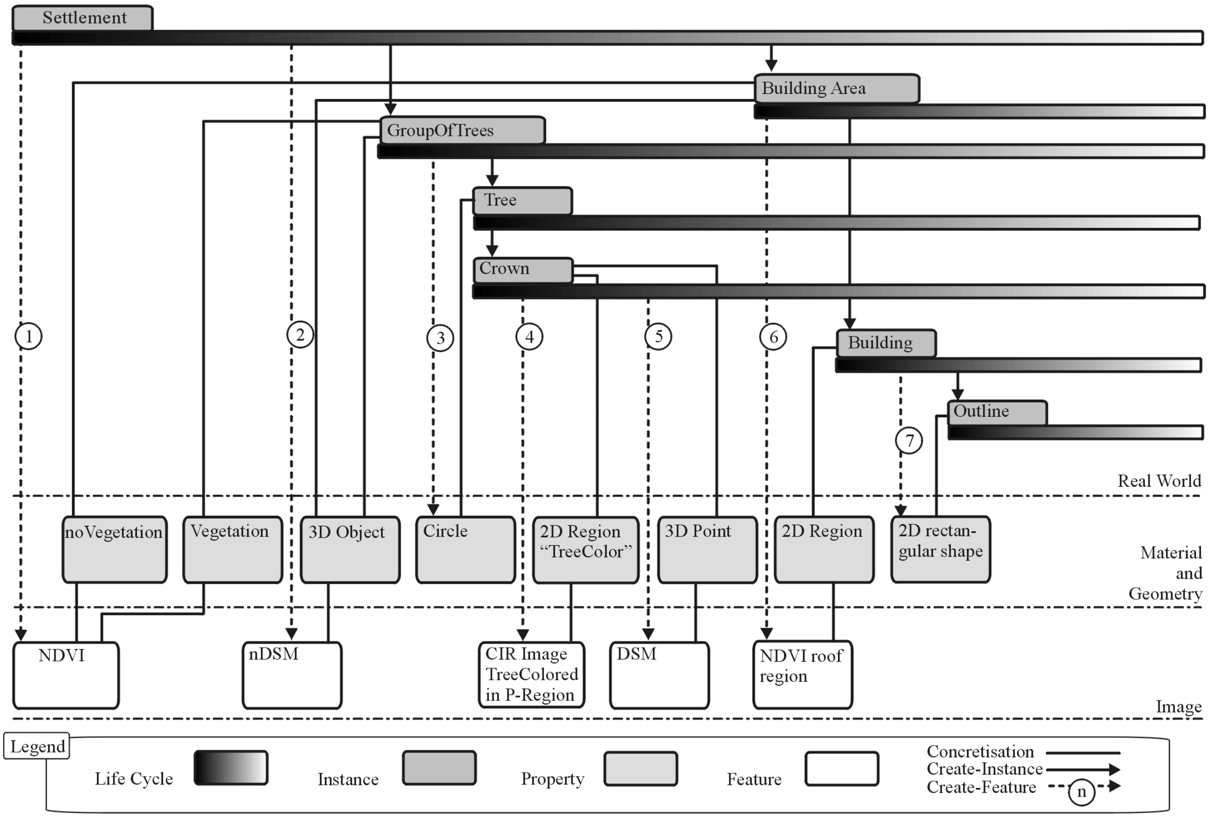

Figure 2: Model and sequence diagram for the extraction of objects in settlement context

We consider a *normalized* DSM (*nDSM*) as input to our approach, which is the difference DSM – DTM. A *digital surface model* (DSM) is an image consisting of height values including vegetation, buildings and other objects. A *digital terrain model* (DTM) consists of only those points lying directly on the terrain. There are various methods to reduce a DSM to a nDSM, in general these algorithms are used for the processing of laser data to derive a DTM (e. g. KRAUS & PFEIFER, 1998, VOSSELMAN, 2000, LOHMANN ET AL., 2000). A discussion of these techniques is beyond the scope of this paper. In order to actually produce a nDSM for our work we have adopted the method described by JACOBSEN & PASSINI (2001) in which the transformation from DSM to nDSM involves a local analysis of the height differences of neighboring points and profiles followed by linear prediction.

4.1 Detection of *Areas*

We use the term *3DObject* for the objects in the scene having a height above the terrain, for example a building in contrast to a road. In the nDSM *3DObjects* are searched for as local maxima having a minimum size. In order to detect vegetation in the scene the NDVI image is segmented into the two regions vegetation and non-vegetation. This segmentation is based on a histogram analysis of the NDVI image under the assumption that there exists a clear local minimum close to zero which represents the threshold necessary for the segmentation. After creation of the necessary features (marked as (1) and (2) in figure 2), instances of the concept *GroupOfTrees* are generated by intersecting the *Vegetation* areas with the detected *3DObjects*. Instances of *BuildingAreas* are generated by intersecting *noVegetation* areas with *3DObjects*.

4.2 Extraction of Trees

The extraction of trees is based on a model, which includes geometric and radiometric features, as well as neighborhood relations. The visible part of a tree in the scene, the crown, is generalized to a circle. Furthermore we assume, that the shape is symmetric around the Z axis. In the radiometric tree model we assume that a tree has a spectral signature different from the signatures of the neighboring trees. This assumption is needed to separate neighboring trees. As in an urban environment the trees are often planted by humans, trees usually stand at some distance. In our model we assume that the distance d between trees is larger than their radius b (see fig. 3).

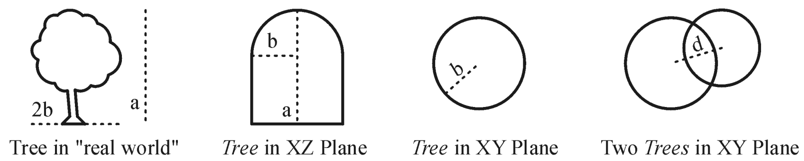

Figure 3: A tree in the "real world" and the geometric parts of the concept *Tree*

After the instantiation of a *GroupOfTree* object the extraction of individual trees is performed in three steps. (1) every *GroupOfTree* region is decomposed into circular regions by means of morphological operations, (2) inside of every circular region the spectral signature is trained and used for a refinement of the position of the radius, and (3) the height of every tree is computed with a kind of a stamp. These steps are mapped in figure 2 as (3), (4), and (5). A detailed description of the algorithm for the extraction of trees is given in STRAUB & HEIPKE (2001).

4.3 Building Extraction

Before analyzing the *BuildingArea* in order to reconstruct the building shape a second segmentation of this region has to be performed. This is motivated by the fact that the global segmentation process can't properly separate individual buildings from its surrounding area caused by a weak resoluted nDSM. To reach a better delineation of the *BuildingArea* a local investigation of the NDVI is carried out: Assuming that the spectral characteristic of the roof is homogenous inside the roof area but differs significantly from the surrounding a minimum and a maximum threshold are selected from a central region in the domain of the *BuildingArea* in the NDVI image, assigned with (6) in fig. 2. The regions resulting after applying these thresholds will now be further described by the analysis of invariant moments.

Invariant moments up to the second order (HARALICK & SHAPIRO, 1993) can be used to define a rectangle around a cloud of points. The parameters of the rectangles -width, length, orientation and position- are derived from analyzing invariant moments. In general the geometric moment M of the order (i,j) in the discrete domain can be written as

$$M_{ij} = \sum_{x_1}^{x_n} \sum_{y_1}^{y_n} x^i y^j f(x,y), \tag{1}$$

where $f(x,y)$ in case of a nDSM is the height H at (x,y). The nDSM acts as a weighting function. The analysis of geometric moments with the aim to reconstruct buildings from height data was proposed by MAAS (1999). The procedure can be summarized as follows: The geometric moments can be transformed to invariant moments by (1) relating the coordinates to the center of gravity (shift invariance) and (2) by means of a principal axis transformation (rotation invariance). Because these invariant moments refer to the local coordinate system of the point cloud, they can be used to calculate the dimension in the local x- and y-direction. The calculation of the orientation gets numerical instable in cases where the shape of the point cloud gets quadratic or round. Additionally when this shape is very complex the orientation may be inaccurate.

The main idea in our approach is that a building shape can be decomposed into more than one rectangle. The principal task is to minimize the difference between the area the point cloud derived from intersecting the *BuildingArea* with the nDSM covers in the XY-plane and the area of the rectangular model. In order to avoid a too fine decomposition a minimum size criteria is brought in.

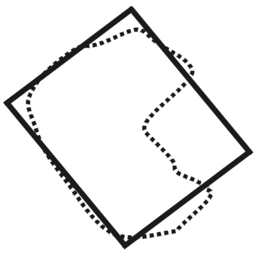

Figure 4: First rectangular describing the whole segmented region

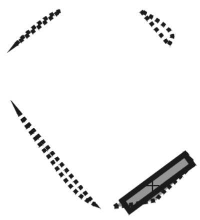

Figure 5: Areas which are not covered by the rectangular model

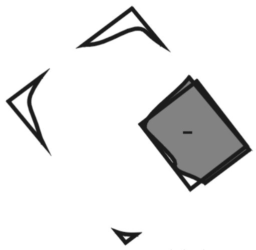

Figure 6: Areas which are not part of the building area, but part of the rectangular model

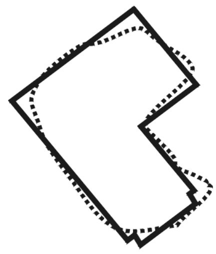

Figure 7: Rectangular model after the first iteration step

In figures 4 to 7 the single iteration steps of our hierarchical approach are presented. In fig. 4 the resulting rectangle after the analysis of invariant moments of the whole region is shown. The dotted line represents the region stemming from the local segmentation, and the black line represents the rectangle, which was derived by analyzing this region. The next step consists in studying the parts of the region which are contained in the region coming from segmentation but not in the rectangle (fig. 5) and vice versa (fig. 6). The regions larger than the minimum size are filled. Figure 7 represents the situation when the accepted regions were described by rectangles and added resp. removed to/from the primary rectangle.

Afterwards the process restarts with the new shape until no remaining areas fulfill the minimum size criteria. In fig. 2 this step can be found as (7). This kind of decomposition allows to gain more geometrical knowledge of the building: Extensions of buildings are added successively to the model and can therefore later on be treated separately (e.g. for the 3D-reconstruction). In addition the use of weighted moments, i.e. taking in account the height, allows to bring in constraints regarding the roof type, e.g an extension of a building may just be added to the description if the roof inclination meets certain size criteria.

5 Results

The approach was applied to image and height data of a test area in Grangemouth, Scotland. The color infrared aerial images were acquired in summer 2000 for the IST project CROSSES. 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 21µm. 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 km². A large part of the whole test site belongs to an industrial plant with sparse vegetation. We have selected a subset of the data set for our test with relatively typical suburban characteristics. One family houses, some larger buildings, trees and roads are visible in this subset. It is about 2700*2300 pixels large, corresponding to an area of approximately 60,000 m², refer to figure 8 for an overview of the "Grangemouth" test site.

Figure 8: Overview of the test site "Grangemouth", two enlarged subsets showing a building and trees in the orthoimage

Figure 9: Enlarged subsets from the test area, *GroupOfTrees* and individual *Trees* superimposed (left), *BuildingArea* before and after local segmentation, final rectangular description (right)

For the two examples depicted in figure 9 the small image on the left shows the surrounding polygon of a *GroupOfTrees* object with the corresponding image. The larger image shows the resulting individual *Trees* by means of a circle superimposed to the image. The right side of figure 9 shows the example building: In the upper small picture the initial *BuildingArea* is superimposed to the NDVI image. The area contains shadily regions, after the local segmentation (ref. chapter 4.3) the area better fits to the building shape (lower small picture). The right image shows the rectangular model of the building superimposed to the orthoimage.

Overall, 235 trees and 46 buildings were extracted in the test site with the approach described above. In order to estimate the *completeness* (True Positives / (True Positives + False Negatives)) and the *correctness* (True Positives / (True Positives+False Positives)) we have compared the obtained results to reference data. A completeness of 95% for the trees, 94% for the buildings and a correctness of 89% for the trees and 80% for the buildings was reached, refer for example WIEDEMANN ET AL. (1998) regarding the quality measures.

6 Summary and Outlook

We have presented an approach for the extraction of trees and building shapes from color infrared and height data in an urban environment for the generation of city models. Regarding trees, the approach is based on a model, using geometric and radiometric features as well as neighboring relations between trees. Building shapes are approximated using the analysis of invariant moments. We have embedded this tree and building shape extraction into a hierarchical scene description of a landscape in different abstraction levels. The abstraction levels are used as a control mechanism for the feature extraction in order to simplify the automatic interpretation of images. The approach was investigated in a test site of approximately 60,000 m^2 with promising results.

The internal evaluation of the results will be the main focus of our work in the future, because it is seen as the most critical point in the proposed automatic system. This is for example important for the orientation of buildings, therefore we intend to enhance the calculation of this parameter. Another idea is to verify the reconstructed corners in the orthoimage.

7 Acknowledgement

This work was developed within the IST Program CROSSES financed by the European Commission under the project number IST-1999-10510.

8 References

BRANDTBERG, T. & WALTER, F., 1998: Automated delineation of individual tree crowns in high spatial resolution aerial images by multiple scale analysis, Machine Vision and Applications 1998 (**11**): 64-73.

BRENNER, C. & HAALA, N., 1999: Extraction of buildings and trees in urban environments, ISPRS Journal of Photogrammetry and Remote Sensing 54 (1999): 130-137.

BRUNN, A., WEIDNER, U., FÖRSTNER, W., 1995: Model-based 2D-Shape Recovery, in: Mustererkennung 1995, G. Sagerer et al. (eds.): 260-268, Springer-Verlag.

BRUNN, A. & WEIDNER, U., 1997: Extracting buildings from digital surface-models. In IAPRS, Vol. 32, Part 4-4W2, “3D Reconstruction and Modeling of Topographic Objects”, 27-34.

CROSSES, 2001: The CROSSES Homepage. http://crosses.matrasi-tls.fr/. (June-01- 2001)

GABET, L., GIRAUDON, G., RENOUARD, L., 1994: Construction automatique de modèles num'eriques de terrain haute résolution en milieu urbain. Société Francaise de Photogrammétrie et Télédétection (135): 9-25.

HARALICK, R.-M. & SHAPIRO, L. G., 1993: Computer and Robot Vision, Vol. II, Addison Wesley, 630 pages.

HEIPKE, C., PAKZAD, K., STRAUB, B.-M., 2000: Image Analysis for GIS Data Acquisition. Photogrammetric Record 16(**96**): 963-985.

JACOBSEN, K., PASSINI R., 2000: Filtering of digital elevation models. In GIS 2001 "Branching Out: Spatial Technology Goes Mainstream", Vancouver, Canada. 19-22 February 2001. (on CD).

KRAUS, K., PFEIFER, N., 1998: Determination of terrain models in wooded areas with airborne laser scanner data. ISPRS Journal of Photogrammetry & Remote Sensing 53 (1998): 193-203.

LINDEBERG, T., 1994: Scale-space theory in computer vision. Kluwer Academic Publishers, Boston, USA. 423 pages.

LOHMANN, P., KOCH A., SCHAEFFER, M., 2000: Approaches to the Filtering of Laser Scanner Data, International Archives of Remote Sensing and Photogrammetry, Vol. 33, Part B 3/1: 534-541.

MAAS, H.G., 1999: Closed solutions for the determination of parametric building models from invariant moments of airborne laserscanner data. International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part B3: 193-199.

POLLOCK, R. J., 1994: A Model-based approach to automatically locating tree crowns in high spatial resolution images, In Desachy, J. (ed.), SPIE Vol. 2315, Image and Signal Processing for Remote Sensing. 526-537.

STRAUB, B.-M., HEIPKE, C., 2001: 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), Birkhäuser, (in Press).

STRAUB, B.-M., WIEDEMANN, C., HEIPKE, C., 2000: Towards the automatic interpretation of images for GIS update, International Archives of Remote Sensing and Photogrammetry, Vol. 32, B2: 521-532.

SUETENS, P., FUA, P., HANSON, A. J., 1992: Computational strategies for object recognition. ACM Computing Surveys 24(**1**): 5–60.

VOSSELMAN, G., 2000: Slope based filtering of laser altimetry data. International Archives of Remote Sensing and Photogrammetry, Vol. 33, Part B 3: 935-942.

WEIDNER, U., 1997: Gebäudeerfassung aus Digitalen Oberflächenmodellen. Deutsche Geodätische Kommission, Series C, 474. 184 pages.

WIEDEMANN, C., HEIPKE, C., MAYER, H, JAMET, O., 1998: Empirical Evaluation of Automatically Extracted Road Axes. In: Bowyer Phillips (eds.), Empirical Evaluation Methods in Computer Vision, Los Alamitos, California, IEEE Computer Society Press. pp. 172-187.