

# Automatic Extraction of Trees and Buildings from Image and Height Data in an Urban Environment

Bernd-M Straub, Markus Gerke, Andreas Koch

Institute for Photogrammetry and Engineering Surveys, University of Hannover, Nienburger Strasse 1  
30167 Hannover, Germany,

straub@ipi.uni-hannover.de, gerke@ipi.uni-hannover.de, koch@ipi.uni-hannover.de,

**Abstract:** The automation of the extraction of objects from aerial and satellite optical images is one of the main research tasks in photogrammetry and computer vision. Our aim in this work is to identify how extraction of trees and buildings from imagery can be improved by the generation of normalized digital surface models from height data, and vice-versa. Aerial images and a hierarchical structured generic model of the real world are used as the only source of information in order to be independent from possibly outdated GIS data.

**Keywords:** Automatic Image Analysis, Building and Vegetation Extraction

## 1 Introduction

In this paper we present ongoing recent work in the domain of the automatic extraction of topographic objects from images and height data. The main focus of our work is on the extraction of buildings and trees in an urban environment. High resolution color infrared (CIR) ortho images with a ground sampling distance (GSD) of 10 cm and a surface model (DSM) with an GSD of 20 cm are used as source of information. The DSM was automatically computed by the French company ISTAR based on the original multiple overlapping image data. One aim of our work is the automatic production of a highly detailed 3D city model, which can be used in a wide variety of applications like games, tourist applications, and simulation computations, refer (CROSSES, 2000). These applications demand on the one hand inexpensive data, on the other hand it is often possible to decrease the data quality requirements. The main section of the paper deals with the model for the extraction of buildings and trees from aerial images and height data. The obtained results are explained by means of an example.

## 2 Description of the Approach

The extraction of objects from images is generally based on a more or less precise model of the objects one wants to detect in the images. In the following we will use the term *scene* as a generic term for all image data, which are used for the extraction of the objects. *Scene description* is used as generic term for all descriptions of objects of the real world which are depicted in the scene. The scene description is often structured hierarchically in different abstraction levels (HEIPKE et al., 2000; STRAUB et al., 2000). For example, roads, buildings and trees are parts of the city, a crown is a part of a tree, and a roof is a part of a building. In our case the city is the highest abstraction level, and crown and roof are concepts of the lowest level of abstraction. Semantic nets are well suited for the description of such object hierarchies. They are often used in topographic applications of image analysis for the formal representation of these models, refer e.g. to (SOWMYA & TRINDER, 2000) for an overview of representation formalism, (NIEMANN et al., 1990) for an early implementation. A semantic net is a directed acyclic graph, whereby the nodes are abstract descriptions of the objects in the scene, and the edges describe the relations between these objects.

On the highest abstraction level the model of an urban environment consists of the concepts *SealedArea*, *VegetationArea*, *3DObjects* and *Terrain*. The concepts are quite simple, a *VegetationArea* is mainly covered by vegetation and a *SealedArea* is mainly sealed. *3DObjects* are objects, which are looked upon as objects with an individual 3-dimensional geometry, like a tree in contrast to a road. Every object which throws shadow is a *3DObject*. The *Terrain* is the description of the terrain without *3DObjects*. The highest abstraction level, assigned as *Coarse Scene Description*, is used for the reduction of the valid domain of the scene, sometimes called *focus of attention*. The next level of abstraction are the concepts called *BuildingArea* and *TreeArea*, which are specializations of the concept *3DObject*. Additionally a *BuildingArea* is sealed and a *TreeArea* is not.

### 2.1 Coarse Scene Description

Instances of the concepts *SealedArea* are regions in the scene with an NDVI value smaller than 0, every region with a NDVI value larger than 0 is labelled as *VegetationArea*. The NDVI is defined as the ratio of the brightnesses  $[\text{red} + \text{nir}]$  and  $[\text{nir} - \text{red}]$  (RICHARDS & JIA, 1999). The instances of the concept *Terrain* are all the regions in the scene which belong to the terrain and not to objects above the terrain like buildings or trees. Theoretically, these regions can easily be computed from the difference of the digital surface model (DSM) and the digital terrain model (DTM), the so called normalized digital surface model (nDSM). All

regions in the nDSM with the value = 0 are instances of the concept *Terrain*. Regions with a value larger than zero lead to instances of the concept *3DObjects*. Instances of the next abstraction level (*BuildingArea* and *TreeArea*) can be derived from these objects.

The normalized DSM is needed for the instantiation of *3DObjects*. The reduction of the surface to the terrain is done by a combination of analyzing height differences of neighboring points, profiles in the direction of row and column and by use of linear prediction (JACOBSEN, 2000). The linear prediction is based on the correlation of neighboring points, expressed in the covariance function. By means of this function a mathematical surface can be calculated, and the differences between this surface and the real height value will be compared with a predefined threshold (LOHMANN et al. 2000, KRAUS & PFEIFER, 1998). The result of these analysis is a DTM, the arising holes are filled by interpolation.

## 2.2 Extraction of Buildings and Trees

As described above, the instances of the concepts *BuildingArea* and *TreeArea* are generated only from the given semantic context information without further image processing. Roof outlines and the crowns of trees are extracted based on the geometric and semantic knowledge stemming from the global scene description.

### 2.2.1 Reconstruction of Building Outlines

The main idea in our approach is that the roof outline can be decomposed into rectangles. The parameters of the rectangles -height, length, orientation and position- are directly derived by analyzing invariant moments. The ratios of invariant moments up to the second order can be used to define a rectangle around a cloud of points (HARALICK & SHAPIRO, 1993). The analysis of geometric moments with the aim to reconstruct buildings from height data was successfully applied by (MAAS, 1999). To "count" and to separate the buildings in the regions coming from the coarse *Scene description* a local investigation of the NDVI is fulfilled: assuming that most of the *BuildingArea* region really contain buildings a minimum and a maximum threshold are selected from a central region in the domain of the *BuildingArea* in the NDVI image. Applying these thresholds to the initial point cloud a second segmentation is performed.

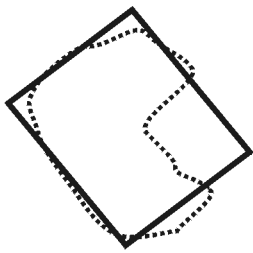


Fig. 1: First rectangle describing the whole segmented region

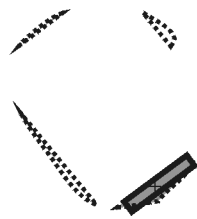


Fig. 2: Areas which are not covered by the rightangular model

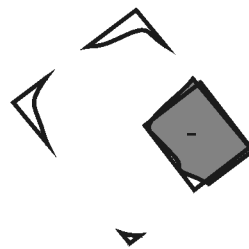


Fig. 3: Areas which are not part of the *BuildingArea* but part of the rightangular model

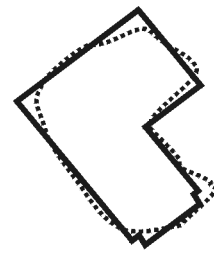


Fig. 4: The rightangular model after the first iteration step

In general, a roof outline cannot be described by a single rectangle. Therefore, the roof outline is assumed to be composed of more than one rectangle. The principal task is to minimize the difference between the area the point cloud 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. In figures 1 to 4 the single iteration steps are presented. In fig. 1 the resulting rectangle after the analysis of invariant moment of the whole region is shown. The dotted line represents the region stemming from the initial segmentation, and the black line represents the rectangle, which was derived by analyzing this region. The next step consists of studying the parts of the region which are contained in the region coming from segmentation but not in the rectangle (fig.2) and vice versa (fig. 3). The regions which fulfill the minimum size criteria are filled. The last part of the figure represents the situation when the accepted regions were described by quadrilaterals and added resp. removed to/from the primary rectangle. Afterwards the process restarts with the new right angled outline until no remaining areas fulfill the minimum size criteria.

### 2.2.2 Extraction of Trees

The extraction of trees is divided into two main tasks, the first is the initialisation of the parameters for every tree by means of the given regions labelled as *TreeArea*, and the second step is the verification and refinement of the parameters in the corresponding domain of the *scene*. The variability of the shapes of trees in the real world is simplified to the following surface of second order.

$$\frac{(z - z_0)^n}{a^n} + \frac{\left( (x - x_0)^n + (y - y_0)^n \right)^{\frac{n}{2}}}{b^n} = 1 \quad (1)$$

This surface was proposed for the approximation of a crown of a tree by (POLLOCK, 1994), whereby  $x$ ,  $y$ , and  $z$  are local co-ordinates,  $b$  is the radius of the crown, and  $a$  is the height above the terrain. The

parameter  $n$  characterizes the shape of the surface, values smaller than 2 lead to a cone, and if  $n$  increases from 2 the resulting shape becomes more and more cylindrical. It was proposed by (MAYR et al., 1999), to use this parameter for the differentiation between coniferous and deciduous trees. We are using the parameterization for the generation of a generic template, which is used to establish the position of a tree in the nDSM in the last refinement step. The parameter  $x_0$ ,  $y_0$ , and  $z_0$ , are transformation parameters to the coordinate system of the scene, i.e. the position of the tree. The radiometric model is mainly based on the assumption that one tree has one characteristic color, which allows to differentiate between two neighboring trees.

The initialization of the tree position and radius is based on the assumption, that in the real world the trees in an urban environment are often placed in that way, that the crowns do not “disturb” each other, when the trees are growing. That means, the mean diameter of two neighbored trees is larger than the distance between these trees. If this assumption is fulfilled, a region *TreeArea* cropped from the normalized digital surface model with high values in the NDVI should have a shape like the schematic one in fig. 5. By means of morphological operations the largest possible circle can be fitted into this region.

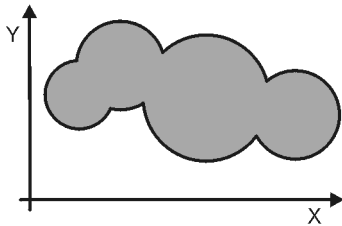


Fig. 5: *TreeArea*

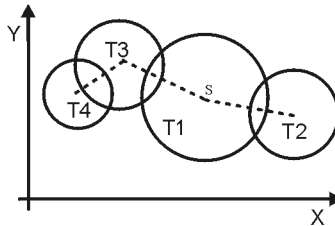


Fig. 6: Decomposed *TreeArea* (Optimal Result)

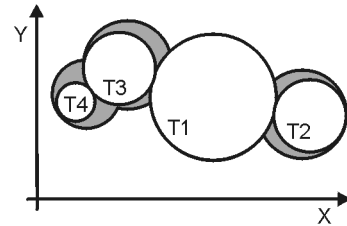


Fig. 7: Decomposed *TreeArea*

This circle is used to initialize the trees with an approximation of the position and radius, afterwards this circle area is removed from the initial area and the next largest circle is fitted to the residual region. The best possible result is shown in fig. 6, four circles, every circle a reliable candidate for a tree. But, one has to take into account, that this is an idealized scheme which can not be reached in praxis, due to the relatively noisy data. The removing of circle  $T(n)$  from *TreeArea* lead to a systematic error in position and radius of the next circle  $T(n+1)$ , refer fig. 7. These error will be reduced in the following step by means of the color information.

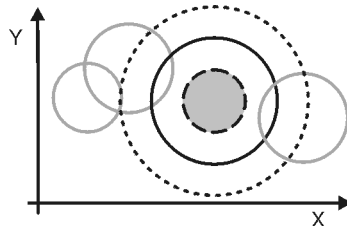


Fig. 8: Regions S, C and P of a Tree

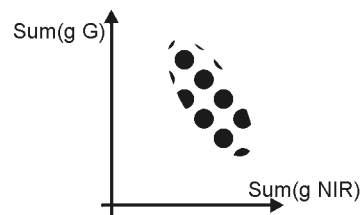


Fig. 9: Feature Space (Region S)

Two additional regions were defined for every tree after the first coarse approximation of the position in  $x_0$  and  $y_0$  - a *safe* region  $S(x_0, y_0, a/2)$ , grey in fig 8, defined by a circle with the half radius at the actual position  $(x_0, y_0)$  used for learning of the characteristic spectral signature of the tree; and a *possible* region  $P(x_0, y_0, a/2)$ , plotted with a dotted line in fig. 8. We assume, that the grey values of the pixels in S are representative for the crown of the tree. These pixels are transformed into the feature space (fig. 9) and the covered domain is marked. All pixel in region P which fit to this domain in the feature space are classified as pixel belonging to the tree. This procedure is quite similar to the strategy a human operator has, if the spectral signature of a specific object class is not normal distributed, refer (ERDAS, 1997). The centre coordinates are computed as the gravity centre of all classified pixels, and the radius is calculated from the area based on the assumption that the area is a circle. At last, P is computed again and a template is generated based on equation (1) with  $n=2$ , both with the refined parameters  $x_0, y_0, b$ . This template is fitted to the DSM, which leads to the missing parameter  $a, z_0$  and the final position  $x_0, y_0$ .

### 3 Results

The approach leads to a scene description including instances of individual buildings and trees. In order to give an idea of the results we show the following examples. Two buildings and four trees are viewable in the subset of the used ortho image, refer fig. 10. The instances of the concepts *BuildingArea* (white line) and *TreeArea* (dotted line) are plotted in fig. 11, with the DSM in the background. The final results are shown in fig. 12. The two buildings in the upper part are extracted mainly correct, the outlines of the roofs are superimposed to the ortho image. In the lower part of the scene, one can see four trees, which were extracted successfully, refer fig. 11. The projection of the crowns onto the XY plane, is represented by the white dotted circles. One can see, that the error in the position which stems from the morphological processing was mainly corrected by means of the refinement of these parameter in the ortho image.



Fig. 10: Ortho Image Used for the Extraction of Trees and Buildings

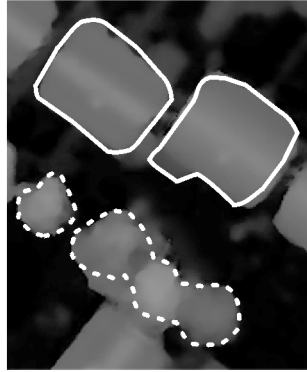


Fig. 11: *BuildingAreas* and *TreeAreas* Superimposed onto the DSM

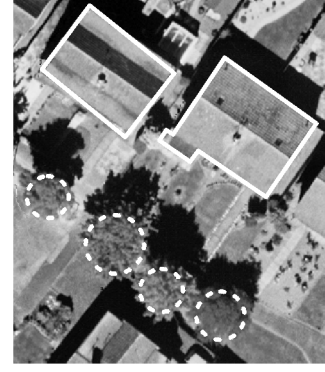


Fig. 12: Extracted Trees and Roof Outlines Superimposed onto the Ortho Image

Summarizing one can say, that the outlines of the extracted objects are a better representation of the 3D objects, than the initial description.

#### 4 Summary and Outlook

The presented approach combines the analysis of height and color infrared image data in order to detect buildings and trees. In the initialization phase the DSM is filtered by means of the linear prediction, the resulting nDSM is used for the generation of hypothesis for 3D objects in the scene. These hypothesis were used for the extraction of trees and buildings from the image and height data. The extracted outlines of the objects are based on a geometric model, which can be used to refine the normalization. Further work will concentrate on the refinement of the building outline by means of snakes and the 3D-reconstruction of buildings using moments of higher order.

#### 5 Acknowledgement

Parts of this work were developed within the IST Project CROSSES financed by the European Commission under the project number IST-1999-10510.

#### 6 Literatur

CROSSES, 2000, CROSSES-Homepage, <http://crosses.matrasi-tls.fr/>, (11.5.2001).  
 ERDAS, 1997, Erdas Imagine V8.3 Tour Guides, Erdas, Inc. Atlanta, Georgia, 454 pages.  
 HEIPKE, C., PAKZAD, K., STRAUB, B.-M., 2000, Image Analysis for GIS Data Acquisition, Photogrammetric Record, 16(96), pp. 963-985.  
 HARALICK, R.-M., SHAPIRO, L. G., 1993, Computer and Robot Vision, Vol.II, Addison Wesley, 630 pages.  
 JACOBSEN, K, 2000, Filtering of digital elevation models, GIS 2000, Vancouver, available on CD.  
 KRAUS, K., PFEIFER, N., 1998. Determination of terrainmodels in wooded areas with airborne laser scanner data. ISPRS Journal of Photogrammetry & Remote Sensing 53 (1998), pp. 193-203.  
 LOHMANN P., KOCH A., SCHAEFFER M., Approaches to the Filtering of Laser Scanner Data, Proceedings, IAPRS, Vol. XXXIII, Part B 3/1, pp. 534-541.  
 MAAS, H. G., 1999, Closed solutions for the determination of parametric building models from invariant moments of airborne laserscanner data. IAPRS, 32, Part 3-2W5, pp. 193-199.  
 MAYR, W., MAYER, H., BACHER, U., EBNER, H., 1999, Automatic Extraction of Trees from Aerial Imagery, in Förstner, W., Liedtke, C.-E. and Bückner, J. (Eds.), 1999. Proceedings, Workshop on Semantic modelling for the acquisition of topographic information from images and maps (SMATI'99), pp. 155-165.  
 NIEMANN, H., SAGERER, G. F., SCHRÖDER, S., KUMMERT, F., 1990, ERNEST: a semantic network system for pattern understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(9), pp. 883-905.  
 POLLOCK, R. J., 1994, Model-based approach to automatically locating tree crowns in high spatial resolution images, SPIE Vol. 2315, Image and Signal Processing for Remote Sensing, Desachy, J. (ed.), pp. 526-537.  
 RICHARDS, J. A., JIA, X., 1999. Remote sensing digital image analysis: an introduction. Third edition. Springer Verlag, Berlin. 363 pages.  
 SOWMYA, A., TRINDER, J., 2000, Modelling and representation issues in automated feature extraction from aerial and satellite images, ISPRS Journal of Photogrammetry and Remote Sensing 55 (2000), pp. 34-47.  
 STRAUB, B.-M., WIEDEMANN C., HEIPKE C., 2000, Towards the automatic interpretation of images for GIS update, IAPRS, Vol. XXXIII, B2, pp. 521-532.