Reference:
M. Gerke, C.Heipke and B.-M. Straub, Building Extraction From Aerial Imagery Using a Generic Scene Model and Invariant Geometric Moments, Proceedings of the IEEE/ISPRS joint Workshop on Remote Sensing and Data Fusion over Urban Areas, November 8-9th 2001, University of Pavia, Rome (Italy), ISBN 0-7803-7059-7, pp. 85-89

Building Extraction From Aerial Imagery Using a Generic Scene Model and Invariant Geometric Moments

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

***Abstract*—The automatic extraction of buildings from aerial imagery in an urban environment is the main focus of this paper. Aerial color infrared images and a digital surface model are used as the source of information. The knowledge about the scene and the geometry of the objects is represented by means of a generic scene model. The strategy of our approach is to reduce the complexity of the image content by means of different abstraction levels. The extraction starts with a description of the coarse content of the given scene. On the scene level the detection of possible building regions is performed. In the next step the knowledge about the surrounding of a building is used in order to support the detection of individual buildings. Finally these buildings are reconstructed using invariant geometric moments leading to orthogonal geometric models.**

Results of our approach are given in the paper. They demonstrate its feasibility and limitations. The practical application background is to provide a detailed semantic and geometric description of an urban environment, useful for a dynamic 3D simulation of a disaster. Our work is embedded in an interdisciplinary research project funded by the European Commission.

***Index Terms*—Image Analysis, Object Detection, Scene Analysis**

I. Introduction

In 1996 the OEEPE (European Organization for Experimental Photogrammetric Research) [1] distributed and analyzed a questionnaire directed to the users and producers of 3D-city models. Beside technical questions the aim of this survey was to get an overview regarding the current and the future demand on the city data. The type of the 55 participating institutions ranged from universities to government agencies and private companies. These participants were classified as data producer or/and data user. Applications mentioned in the survey included architecture, tourist information systems and telecommunications. One of the main objects of interest regarding both production and use were buildings: 95% of the producers and users declared themselves to be concerned with buildings.

Manuscript received October 05, 2001. Parts of this work were developed within the IST Project CROSSES financed by the European Commission under the project number IST-1999-10510.

All authors are with the Institute for Photogrammetry and GeoInformation, University of Hannover, Nienburger Str. 1, D-30167 Hannover, Germany.

M. Gerke (telephone: 0049-511-762-19951, e-mail:gerke@ipi.uni-hannover.de).

In this paper we describe our work on building extraction from aerial imagery in the form of orthoimages and digital surface models. The goal of our work is to generate building description which can be used in a simulation system for training emergency forces [2]. In this regard it is sufficient to extract 2D roof outlines of the buildings and attribute them with a constant height value, resulting in flat roof buildings.

We first give a review on related work. Afterwards our strategy regarding building detection and reconstruction using a generic scene model is introduced. In the last sections results and an outlook are given.

II. Related Work

The process of building extraction from imagery can be separated in two main tasks: first the detection of the object ("Where is a building?"), second the reconstruction ("Which geometrical description can be found for this object?"). Regarding these topics a lot of publications can be found. Some of these deal with the complete process, others concentrate on one of the named tasks. In [3] an extensive overview of the subject can be found. As we will use height models and color-infrared orthoimages with the aim to extract a 2D outline of the roof we will further focus on approaches using similar data and results. The more complex problem of the whole roof-surface-reconstruction as done e.g. in [4] is not addressed in this paper.

Regarding building detection from height models an approach can be found in [5]. Here the detection of buildings is based on a segmentation of a normalized digital surface model, whereas [6] make in addition use of mulitispectral imagery. The spectral information facilitates the distinction between vegetated and non-vegetated areas and supports the segmentation process.

To reconstruct buildings from segmented regions [5] presents two approaches. The 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. [7]) or described by a parametric building model which is at least a rectangle in the 2D-planer described by 5 parameters. On the one side fitting a polygon to the region is more comprehensive than using a parametric model but on the other side the quality of the result depends on a sensitive tuning of control parameters. In [8] invariant geometric moments of the segmented region are

analyzed in order to achieve simple parametric building models. One advantage of using moments is that they directly lead to the five parameters (width, length, orientation and position in x and y) describing a rectangle around the region.

III. Detection of Building Areas

In our work we show how the detection and reconstruction of buildings can be embedded into a generic scene model. The whole process of object extraction can be illustrated in such a model. Regarding the detection of buildings we employ a similar strategy as introduced in [6], because we make also use of color infrared images and height models. Regarding the building reconstruction we extend the approach presented in [8]. The extension is based on a more universal geometric model for buildings.

In recent work on image analysis [9],[10] we have employed an approach using a hierarchical scene model, which leads to a successive reduction of the scene complexity. First the whole scene is subdivided into so called *SuperClasses* (*Settlement*, *OpenLandscape*, *Forest* and *Water*), see also [11]. Concentrating on one of these classes one can take into account context-dependant knowledge for further processing. For example the *Settlement* contains *Areas* such as *BuildingAreas* or *GroupOfTrees*. Following the same pattern one can find objects which are contained in these *Areas*. As an example the part of the generic scene model describing the object *Building* is depicted in fig. 1. The focus in this section is on detecting *BuildingAreas*; for the model regarding *GroupOfTrees* and single *Trees*, refer to [9]. In the scene model the layer below the "Real World"-layer is named "Geometry and Material". This layer describes the physical properties of the objects and is data independent. For example the roof outline of a building is modeled as an orthogonal closed polygon. The bottom layer of the generic scene model is named "Image" layer. The term "image" includes all possible raster data, e.g. optical images, surface models or radar images. In this layer the concretization of the objects as observable in image data is described. For instance a *CastShadow* region doesn't reflect much radiation in the visible spectrum and therefore a "Low Reflectance Area" can be a concretization of such a region. One advantage of this type of object representation is the independence of the single layers: if additional sensors are used their properties can be added to the "Image"-layer without changing any other layers.

In our approach we consider two different image data information as input: (1) a normalized Digital Surface Model (nDSM) and (2) a color infrared orthoimage.

A DSM consists of all height values including buildings, vegetation and other objects, a DTM just contains those points situated on the terrain. The difference of these two models - called nDSM - can be represented as an image which contains all objects above the terrain.

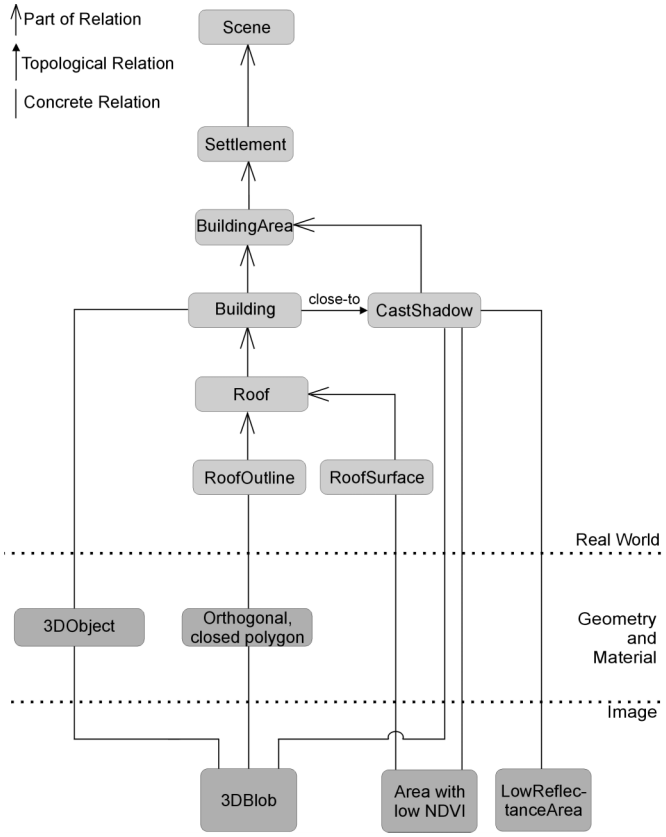

Fig. 1: Generic scene model describing parts of the settlement objects

In the model the *BuildingArea* consists of one or more *Buildings* and close by *CastShadow* regions. The *Building* itself consists of *Walls* and the *Roof*. The *Walls* are not explicitly represented in fig.1, because they do not play a role in our extraction process. The *Building* is a *3DObject* and can therefore be detected in the nDSM. This is expressed in the "Image" layer: a *3DBlob* is a concretization of such a *3DObject*.

The *Roof* has the two parts *RoofOutline* and *RoofSurface*. The first is assumed to be an orthogonal closed polygon. Because of the buildings appearance in the CIR-image the *RoofSurface* is important for its detection. In the generic scene model an area with a low NDVI is a concretization of this concept. In our context a low NDVI is defined as a value in a non-vegetated area. Domains having these two properties (*3DObject* and *NDVI Low*) are said to be instances of the *Building* concept.

On the same abstraction level as the *Building* concept we inserted the *CastShadow* concept; the topological relation between these two objects is "close-to". This is motivated by the fact that 3D objects like buildings cast shadows next to the object. In our model we have introduced a relation between *CastShadow* and *3DBlob*. The reason is that for DSMs generated by digital image matching the resulting *3DBlob* is often enlarged in the direction of the shadow due to the poor matching performance in shadow areas and the subsequent interpolation of the required height values. If DSMs from other sources such as laser scanning are available, the scene model can be somewhat simplified. Due to our definition of a "low

NDVI" *CastShadow* regions are also areas with a low NDVI. The separation of *CastShadow* regions from *Buildings* is discussed in the next section.

IV. Building Extraction

In this section we describe our approach to use the generic model for the extraction of single buildings. The two main steps are (1.) detection of the buildings in the *BuildingArea* and (2.) reconstruction of the single buildings. The image layer in the generic scene model specifies *CastShadow* to be "areas with low reflectance". This property is used in the following. The intensity values are taken from an IHS-transformation of the CIR-image. Provided that only shadow and building information can be found in this area we may assume that the histogram of the *BuildingArea* is bimodal and that the left main peak represents the shadow information whereas the gray values of the right peak belong to one or more buildings (fig. 2). The boundary between these two peaks can be identified by searching for the local minimum after smoothing the histogram. With this threshold the *CastShadow* regions can be removed. According to the model the remaining pixels belong to one or more *Buildings*. The corresponding image regions are further called *Building* regions.

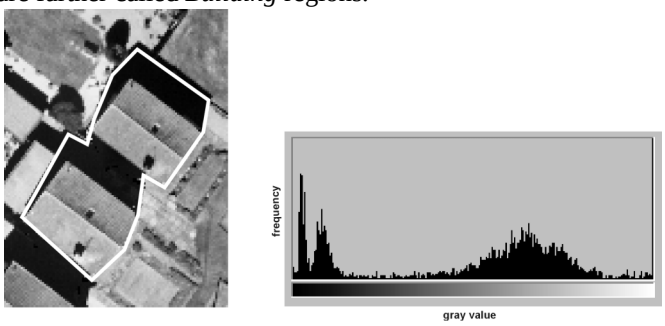

Fig. 2: Exemplary *BuildingArea* and its histogram

The last step is the building reconstruction. The "Geometry" layer contains the concept "orthogonal closed polygon" as concretization of the roof outline. The task is to fit a polygon with these properties into the *Building* region. In computer vision applications moments are often used to derive geometric features of regions [12]. Our approach uses invariant geometric moments as proposed in [8]. However, we extend the described approach by a hierarchical decomposition of the initially obtained results.

In the following some details are given in order to clarify the process. In general the geometric moment *M* of the order (i,j) in the continuous domain can be written as

$$M_{ij} = \int_{x_1}^{x_2} \int_{y_1}^{y_2} x^i y^j f(x,y) dxdy, \tag{1}$$

where *f(x,y)* is a weighting function. As we want to compute measures of the region in 2D we just take into account binarized data, therefore *f(x,y)* is equal to 1. To gain information about properties of the roof one may use the nDSM heights as weight function. Setting *f(x,y)* to 1 and taking into consideration that we are working in the discrete (raster) domain leads to

$$M_{ij} = \sum_{k=x_1}^{x_2} \sum_{l=y_1}^{y_2} {x_k}^i {y_l}^j. \tag{2}$$

In order to derive shift invariance one has to relate the coordinates to the center of gravity. This is

$$\bar{x} = \frac{M_{10}}{M_{00}}, \bar{y} = \frac{M_{01}}{M_{00}}. \tag{3}$$

The shift invariant moments are

$$\overline{M}_{ij} = \sum_{k=x_1}^{x_2} \sum_{l=y_1}^{y_2} (x_k - \bar{x})^i (y_l - \bar{y})^j. \tag{4}$$

The next step consists in obtaining rotation invariance by means of a principle axis transformation. The orientation of the principle axis is

$$\Theta = \frac{1}{2} \arctan \frac{2\overline{M}_{11}}{\overline{M}_{20} - \overline{M}_{02}} \tag{5}$$

and the shift and rotation invariant moments are

$$M_{pq}' = \sum_{r=0}^{p} \sum_{s=0}^{q} (-1)^{q-s} \cdot \binom{p}{r} \cdot \binom{q}{s} \cdot (\cos\Theta)^{p-r+s} \cdot (\sin\Theta)^{q+r-s} \cdot \overline{M}_{(p+q-r-s)(r+s)}. \tag{6}$$

Because these invariant moments refer to the local coordinate system of the region, they can be used to calculate the dimensions of the rectangle *Lx* and *Ly* in the local x- and y-direction: the relevant integrals in (1) can be resolved into:

$$\begin{aligned} M_{00}' &= \int_{-Lx/2}^{Lx/2} \int_{-Ly/2}^{Ly/2} dxdy = Lx \cdot Ly \\ M_{02}' &= \int_{-Lx/2}^{Lx/2} \int_{-Ly/2}^{Ly/2} y^2 dxdy = \frac{1}{12} Lx \cdot Ly^3 \\ M_{20}' &= \int_{-Lx/2}^{Lx/2} \int_{-Ly/2}^{Ly/2} x^2 dxdy = \frac{1}{12} Lx^3 \cdot Ly \ . \end{aligned} \tag{7}$$

This leads to *Lx* and *Ly*:

$$Lx = \sqrt{\frac{12 \cdot M_{20}'}{M_{00}'}}, Ly = \sqrt{\frac{12 \cdot M_{02}'}{M_{00}'}}. \tag{8}$$

The five parameters describing a rectangle around a region can be computed with (3), (5), (8). We call the rectangle around the whole *Building* region the "initial" rectangle.

More complex buildings can not be described sufficiently by a single rectangle. For instance the initial rectangle of a L-shaped building covers areas which do not belong to this

building.

Therefore, in the next step, called decomposition, we search for difference areas not belonging to the *Building* region but belonging to the initial rectangle and vice versa. These areas are fitted to rectangles also, which then are subtracted from the initial rectangle or added to it. The applied method has been described in detail in [10]. In the following we want to discuss a problem mentioned in this publication: the reconstruction of a nearly quadratic building. For such a building the orientation of the rectangle cannot be computed reliably because the denominator in (5) tends towards zero. A miss-oriented initial rectangle leads to a wrong reconstruction of the whole building. Therefore, the orientation of the initial rectangle has to be corrected.

The improvement of the orientation is done as follows: if it is assumed that the extents of two rectangles are identical and if the center of gravity of both figures are identical, too (these assumptions are valid if invariant geometric moments are used), the orientation will be equal if the area covered by both figures is a maximum, see fig. 3 for an example. In the left part the two rectangles are not oriented in the same direction; the common area (in gray) is smaller than in the right part where the two figures have the same orientation. This model is applied in the reconstruction phase: the initial rectangle derived by the analysis of invariant geometric moments is rotated until the area, covered by this rectangle, and the *Building* region becomes a maximum.

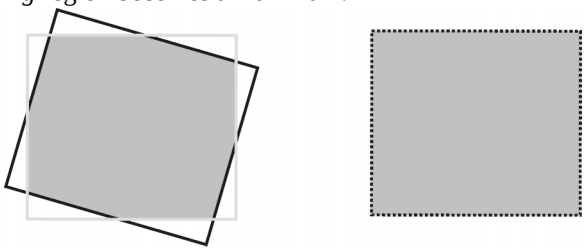

Fig. 3: Model for orientation optimization: The common area of both figures should be maximized

V. Results

The described approach has been applied to the data produced within the framework of the CROSSES-project [2] in which we are participating. The aerial CIR-images have a ground sampling distance (GSD) of 10cm. The along- and cross-track overlap of the photos is 80%. From the images a DSM with a resolution of 20cm has been obtained through image matching [13], and an orthoimage has been also computed. The testsite is situated in Grangemouth, Scotland. In fig. 4 the orthoimage is shown with detected *BuildingAreas* and additional regions which are incorrectly assigned as *BuildingAreas*. These regions are very small, non-compact. or lie at the image border and have been automatically discarded from further processing using heuristic thresholds. From the 97 regions 46 have been accepted as valid *BuildingAreas* and have therefore been further investigated.

Fig. 4: Detected *BuildingAreas* with additional incorrectly assigned regions

The 46 valid *BuildingAreas* contained 55 *Building* regions which have first been separated from the surrounding shadow and afterwards they have been reconstructed (fig. 5).

Fig. 5: Reconstructed buildings in the testsite

In the following some detailed results are shown. In fig. 6, left part, a *BuildingArea* is shown whereas the picture in the middle shows the region after the cast shadow has been removed (*Building* region). In the image on the right side the reconstructed roof outline is shown. One can see that the decomposition works fine in this example.

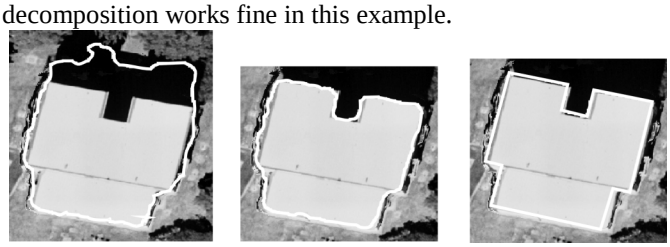

Fig. 6: *BuildingArea* – *Building* region – reconstructed building with orientation optimization enabled

The next figure (fig. 7) shows the same building like in fig. 6, but with disabled orientation optimization. One can see that the initial rectangle is miss-oriented and therefore the following decomposition leads to wrong results.

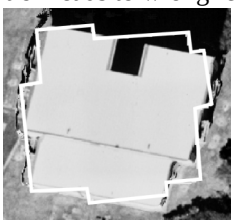

Fig. 7: Same building like in fig. 6, but with disabled orientation optimization

In fig. 8 a similar building is shown. Here the cast shadow removal process did not separate the shadow sufficiently from the *Building* region. The consequence is that the orientation optimization leads to a miss oriented initial rectangle. This example shows that it is very important to put emphasize on the initial detection of the *Building* region.

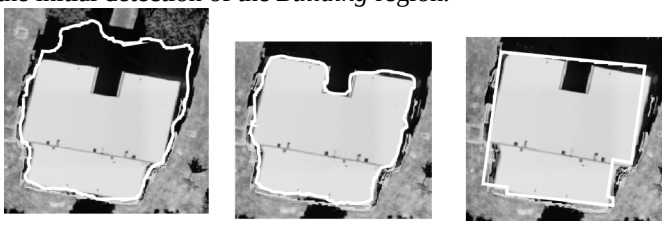

Fig. 8: : *BuildingArea – Building* region– reconstructed building with orientation optimization enabled

In the last figure (fig. 9) a simulated L-shaped *Building* region is shown in order to once again demonstrate the advantages of the decomposition. This L-shaped "building" is not exactly orthogonal, but according to the geometric model the reconstruction procedure approximated an orthogonal polygon.

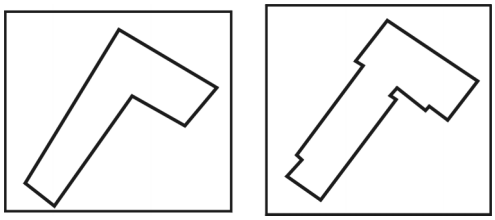

Fig. 9: Simulated *Building* region– reconstructed building

VI. Summary and Outlook

This paper continues our former work on object extraction in urban environments. We embedded the building extraction into a framework of a generic scene model. The approach uses a height model and the NDVI in order to detect *BuildingAreas*. According to the scene model in these areas *CastShadow* regions and one or more *Building* regions are situated. In order to separate the single *Building* regions an investigation of the gray value histogram of the *BuildlingArea* is performed. Afterwards the obtained *Building* regions are fitted to a orthogonal polygon. This geometric model is also part of the scene model. The fitting is carried out using invariant geometric moments which successively decompose the given region into rectangles.

In the near future we will extend our work in the following areas: use of information (edges, corners) in order to improve the orientation of the roof outline; combined extraction of buildings, roads and tree objects and validation of the obtained results by comparing them to independently captured reference data.

References

[1] C. Fuchs, E. Gülch, and W. Förstner, "OEEPE Survey on 3D-City Models," *OEEPE Publication N° 35*. Bundesamt für Kartographie und Geodäsie. Frankfurt. 1998, pp. 9-123.

[2] CROSSES, *The CROSSES Homepage*. http://crosses.matrasi-tls.fr/. (September-01- 2001).

[3] H. Mayer, "Automatic Object Extraction forom Aerial Imagery – A Survey Focusing on Buildings," *Computer Vision and Image Understanding*, vol. 74, 1999, pp. 138-149.

[4] C. Brenner, "Towards fully automatic generation of city models," *International Archives of Photogrammetry and Remote Sensing*, vol. XXXIII (B3), Amsterdam, 2000, pp. 85-92.

[5] U. Weidner, "Gebäudeerfassung aus Digitalen Oberflächenmodellen," Deutsche Geodätische Kommission, Series C, 1997, 474. 184 pages.

[6] C. Brenner and N. Haala, "Extraction of buildings and trees in urban environments," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol 54, 1999, pp 130-137.

[7] I. Rissanen, "Minimum Description Length Principle," *Encyclopedia of Statistical Sciences*, vol. 5, 1987, pp. 523-527.

[8] H.-G. Maas, "Closed solutions for the determination of parametric building models from invariant moments of airborne laserscanner data," *International Archives of Photogrammetry and Remote Sensing*, vol. XXXII (B3),1999, pp. 193-199.

[9] B.-M. Straub and C. Heipke, "Automatic Extraction of Trees for 3D-City Models from Images and Height Data,"*Automatic Extraction of Man-Made Objects from Aerial and Space Images*, vol. III, Birkhäuser, 2001, in press.

[10] M. Gerke, B.-M. Straub and A. Koch, "Automatric Detection of Buildings and Trees from Aerial Imagery Using Different Levels of Abstraction," *Publications of the German Society for Photogrammetry and Remote Sensing"*, vol X, E. Seyfert, Ed., 2001, pp. 273-280.

[11] R. Tönjes, "Wissensbasierte Interpretation und 3D-Rekonstruktion von Landschaftsszenen aus Luftbildern," *Fortschrittberichte VDI*, vol. 10, 575, 1998.

[12] R.M. Haralick and L.G. Shapiro, *Computer and Robot Vision*, vol. II, 1993, 630 pages, Addison Wesley.

[13] L. Gabet, G. Giraudon and L. Renouard, "Construction automatique de modèles num'eriques de terrain haute résolution en milieu urbain, " *Société Francaise de Photogrammétrie et Télédétection*, vol 135, 1994, pp. 9-25.