ISPRS Journal of Photogrammetry & Remote Sensing 54 (1999) 130–137

PHOTOGRAMMETRY & REMOTE SENSING

Extraction of buildings and trees in urban environments

Norbert Haala *, Claus Brenner

Institute for Photogrammetry, University of Stuttgart Geschwister-Scholl-Strasse 24, D-70174 Stuttgart, Germany

Received 25 September 1998; accepted 25 February 1999

Abstract

In this article, two methods for data collection in urban environments are presented. The first method combines multispectral imagery and laser altimeter data in an integrated classification for the extraction of buildings, trees and grass-covered areas. The second approach uses laser data and 2D ground plan information to obtain 3D reconstructions of buildings.

Keywords: virtual city models; 3D modeling; classification; building extraction; tree extraction

1. Introduction

Airborne laser scanners measure three-dimensional points, which are distributed over the terrain surface and on objects rising from the ground like trees or buildings. Even though a so-called Digital Surface Model (DSM), which can be derived from these points is easily sufficient for a number of applications, further interpretation and qualification of the original height data is necessary in many cases. One example is the acquisition of 3D city models for virtual reality applications, where even small geometric errors like a non-planar triangulation of a planar facade may easily disturb the impression of looking at a ‘real world’ scene.

Another problem with employing surface descriptions by an unqualified, i.e., object independent, distribution of points is the large amount of data to be computed, stored and displayed during simulations and visualizations. It has been recognized early by the computer graphics community that handling of very large scenes is an important topic, resulting in several approaches like clipping and viewing frustum computations, if just parts of the scene need to be rendered for visualization. Another traditional approach is to provide multiple level-of-detail representations, which are computed off-line and use different object representations depending on the distance of the (virtual) camera from the object. Although the amount of data can be reduced significantly by these algorithms, many tasks aiming at visualizations or simulations in an urban environment need further abstraction and interpretation of the surface description. Simulations on the propagation of noise or electro-magnetic waves, e.g., require knowledge about the surface material; consequently, trees or buildings have to be represented separately from the terrain surface. Virtual reality applications require terrestrial images to be mapped onto the vertical faces of the buildings to achieve photorealism for the generation of walk-throughs, since the

* Corresponding author. Tel.: +49-711-121-3383; Fax: +49-711-121-3297; E-mail: norbert.haala@ifp.uni-stuttgart.de

0924-2716/99/$ - see front matter
PII: S0924-2716(99)00010-6

resolution and viewing direction of an aerial image is no longer sufficient at a large scale. For this purpose, correspondences have to be established between the terrestrial images and the data set used to represent building geometry. Corners and edges of buildings can be identified easily in terrestrial images. Therefore, these elements should also be represented explicitly in the geometric database. All those arguments call for a further interpretation and abstraction of the laser data in order to achieve an explicit representation of building geometry and topology by 3D CAD models.

Frequently, ground plans of buildings have already been acquired and can therefore be used for 3D building reconstruction. One problem arising when utilizing existing databases is the potential occurrence of errors as well as their prospective lack of actuality and incompleteness. For this reason, the algorithm described in the following section aims at the automatic extraction of topographic features by combining laser data and color imagery in a classification step. Within this step, regions representing buildings and streets are detected, which can be used for map revision to uncover obsolete or incomplete parts of an existing GIS. Additionally, urban vegetation like trees and bushes are captured. These objects are usually not available from standard databases. Still they are very relevant for virtual reality applications, since they play an important role for the visual impression.

The generation of virtual city models including a 3D building reconstruction based on the combination of laser data and existing outlines is described in Section 3. Based on the given ground plan, each building is subdivided into a number of basic primitives and the primitives are fit to the laser data by a least squares estimation. The virtual city model is completed by texture mapping of terrestrial images.

2. Classification using DSM and color imagery

DSM have already been used for the automatic detection of buildings to trigger the subsequent geometric reconstruction from stereo image data (Haala, 1994). This is possible since DSM not only represent the terrain surface like Digital Terrain Models (DTM), but also contain buildings and other objects which are higher than their surroundings. To make this information accessible the so-called normalized DSM, i.e., the difference between DSM and DTM can be calculated, resulting in a representation of objects rising from the terrain approximately put on a plane (Weidner and Förstner, 1995). The required DTM can be provided from existing data or can be derived from the available DSM, for example by mathematical morphology. Even though objects rising from the terrain can be detected quite well from the height data, the discrimination between buildings and trees can be difficult, if only simple criteria like region size or shape are considered. A possible approach is to use the roughness of the DSM surface measured by differential geometric quantities as an additional criterion for the discrimination of buildings and vegetation (Brunn and Weidner, 1997). However, due to the restriction to surface geometry, the number of object types, which can be discriminated within a DSM is very limited. A further differentiation like the extraction of streets or landuse classes is not possible.

Spectral information has been widely used as a data source for thematic mapping applications. Surface material information can be derived by traditional classification techniques from multispectral imagery and used for mapping of man-made structures and natural features in complex urban scenes. Still the potential of spectral data is limited for these applications with respect to the accuracy and reliability of the results as well as the possibility to discriminate a large number of object categories. One problem while classifying multispectral data is the similar reflectance of trees and grass-covered areas. Frequently, the same holds true for the distinction of streets and buildings. On the other hand, trees and buildings can be discriminated easily from grass-covered areas or streets using height data, since they are higher than their surroundings, whereas streets and grass-covered areas are at the terrain level. For this reason, the height above terrain as an additional source of information improves the discrimination and avoids misclassification especially for these types of objects; multispectral images and normalized DSM can be used as complementary data sources to improve the classification.

Hug (1997) applies a scanning laser altimeter, which is able to measure distance *and* surface re-

Fig. 1. CIR orthoimage (green channel selected for grey value representation).

flectance by processing the return signal energy of the laser beam. Thus, in addition to the range data a reflectance image in the near infrared spectrum is available, enabling the simultaneous usage of geometric and radiometric information for the detection of trees and buildings. Henricson et al. (1996) use information from colored infrared aerial images to separate elevation blobs detected in a DSM from stereo image matching into the classes buildings and trees. The application of traditional spectral classification techniques to derive surface material information from multispectral imagery is presented by Ford et al. (1997). They use hyper-spectral data with nominal 2 m ground sample distance with over 200 spectral samples per pixel, captured by the airborne sensor system Hyperspectral Digital Imagery Collection Experiment (HYDICE).

In our approach, multispectral information provided by a color-infrared aerial image is combined with geometric information from a laser scanner DSM in one classification step. The utilized CIR images of the test site Karlsruhe were taken at a scale of 1:5000 with a normal-angle camera. For digitization, the images were scanned at a resolution of 60 μm, resulting in three digital images in the spectral bands near infrared, red and green at a pixel footprint of 30 cm. The basic idea of the proposed algorithm is to simultaneously use geometric and radiometric information by applying a pixel-based classification, whereby the normalized DSM is used as an additional channel in combination with the three spectral bands. For this purpose, the height data and the images have to be coregistrated, i.e., a colored orthoimage is generated from the original CIR imagery. Fig. 1 shows the green channel of the orthoimage, generated with a ground pixel size of 0.5 m. Fig. 2 shows the corresponding DSM, which was provided by the TopoSys laser scanner-system; terrain points were measured at approximately one point each 1×1 m^2 with an accuracy of 0.2 m.

Based on this DSM, the terrain surface can be derived approximately by mathematical morphology using the approach described by Weidner and Förstner (1995). This DTM can then be subtracted from the original DSM. Together with the spectral bands the resulting normalized DSM is used as input for an unsupervised classification algorithm. The applied classification detects clusters of pixels in feature space and categorizes the pixels to the clusters based on the minimum distance criterion. For that purpose the ISODATA (Iterative Self-Organizing Data Analysis Technique) algorithm was utilized, which is a standard procedure (see e.g., Richards, 1993). With this approach, the optimal number of spectral clusters is automatically determined by itera-

Fig. 2. DSM acquired by laser scanning (mosaic of two parts, upper and lower, stemming from different datasets).

Fig. 3. Result of ISODATA classification using CIR orthoimage and normalized DSM.

tively applying split and merge operations while performing the following steps.

(1) A number of parameters have to be initialized by the operator. These parameters are the desired number of clusters, the minimum number of iterations, the minimum number of pixels in a cluster, the minimum standard deviation to initiate cluster splitting, and the maximum distance in feature space between cluster centers to initiate cluster merging.

(2) Each pixel is assigned to one of the predefined clusters by a minimum distance criterion.

(3) All clusters containing less members than a predefined number of pixels are eliminated.

(4) New cluster centers are computed from the pixels assigned in step 2.

(5) Aggregated clusters are split, if the standard deviation is larger than the specified threshold. Neighboring clusters are merged, if the pairwise

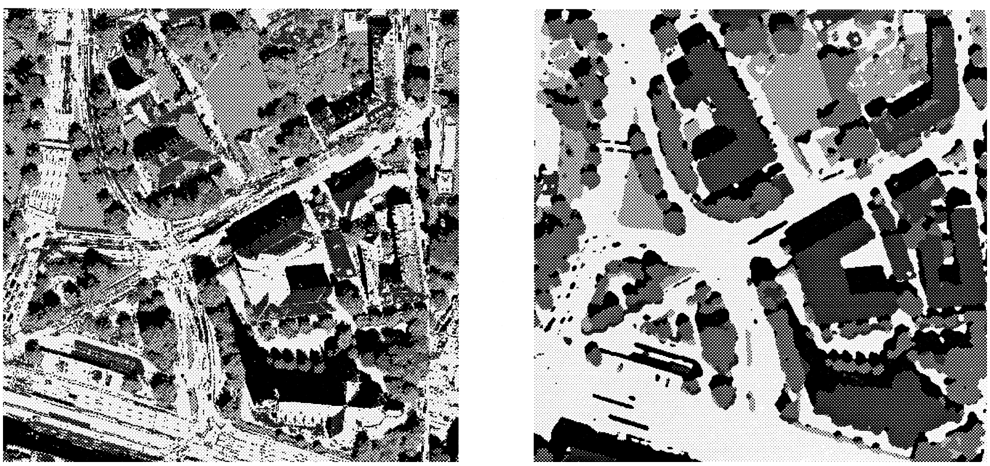

Fig. 4. Comparison of multispectral classification with (left) and without (right) normalized DSM.

distance is smaller than a predefined parameter and if the maximum specified number of clusters has not been reached.

(6) The algorithm is terminated, if the maximum number of iterations is reached, else it is continued with step 2.

For computational reasons only a certain percentage of pixels (for example only the pixels of each 10th row and column) are used for cluster formation in step 2. After the stop criterion is reached, the detected cluster centers are used to classify the entire image based on the minimum distance criterion. Finally, the feature classes are interpreted and combined to thematic classes or object classes in an interactive step.

The results of the ISODATA algorithm of course depend on the quality of the input data. If the data in feature space is distributed in almost isolated natural groups, these clusters can be detected very reliably. Fig. 3 shows the result of the classification based on CIR orthoimage and normalized DSM, where classes (from dark to bright) ‘shadow’, ‘building’, ‘tree’, ‘grass-covered-area’ and ‘street’ could be obtained. In order to demonstrate the importance of the normalized DSM, Fig. 4 compares the result with a classification solely based on spectral information.

This classification can, e.g., be used to automatically determine the position of trees as shown in Fig. 5. In order to define the position of single trees within the class ‘tree’ regions, depicted by the white contours in Fig. 5 a skeleton was calculated for these areas (black lines). The trees were placed at the node points of the skeleton represented by the white points. Fig. 6 shows the generated trees in a virtual environment. The automatic reconstruction of the buildings contained in this figure is described in Section 3.

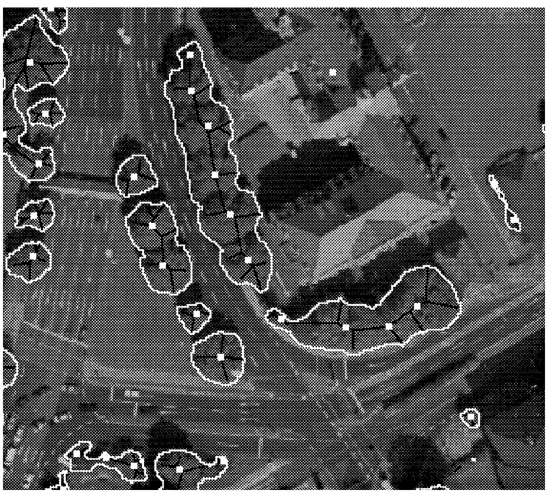

Fig. 5. Computation of tree positions using classified ‘tree’ regions.

The result of the classification can also be used to check existing data for actuality and completeness, e.g., by detecting buildings, which are not present in the database. In principle, the classification can provide a very simple 2.5 D urban model; building blocks can be generated using the outlines of the classified buildings. For more sophisticated tasks, like visualization or simulations of high qualit, a 3D CAD model has to be provided. For that purpose the DSM data is combined with existing ground plans of the buildings, which have been checked previously by the presented classification step.

3. Building reconstruction

Most buildings can be described to sufficient detail in terms of general polyhedra, i.e., their boundaries can be represented by a set of planar surfaces and straight lines. In our approach, additional constraints are obtained for reconstruction by using the assumption that the given ground plan is correct and exactly defines the borders of the roof. The buildings are represented by a combination of one or more basic primitives. Similar to the approach of Englert and Gülch (1996), each primitive consists of a cuboid element with a roof, which can be a flat roof, desk roof (one inclined plane), gable roof or hip roof. This type of representation corresponds to the well-known constructive solid geometry (CSG) representation used in computational geometry, which combines simple primitives by means of Boolean set operators (union, intersection, subtraction) in order to obtain complex objects.

In the first step of our algorithm buildings are segmented into basic primitives based on their given outlines. An example of the automatic ground plan decomposition is shown in Fig. 7. Each rectangle defines the base of one building primitive, which means that position, orientation and horizontal exten-

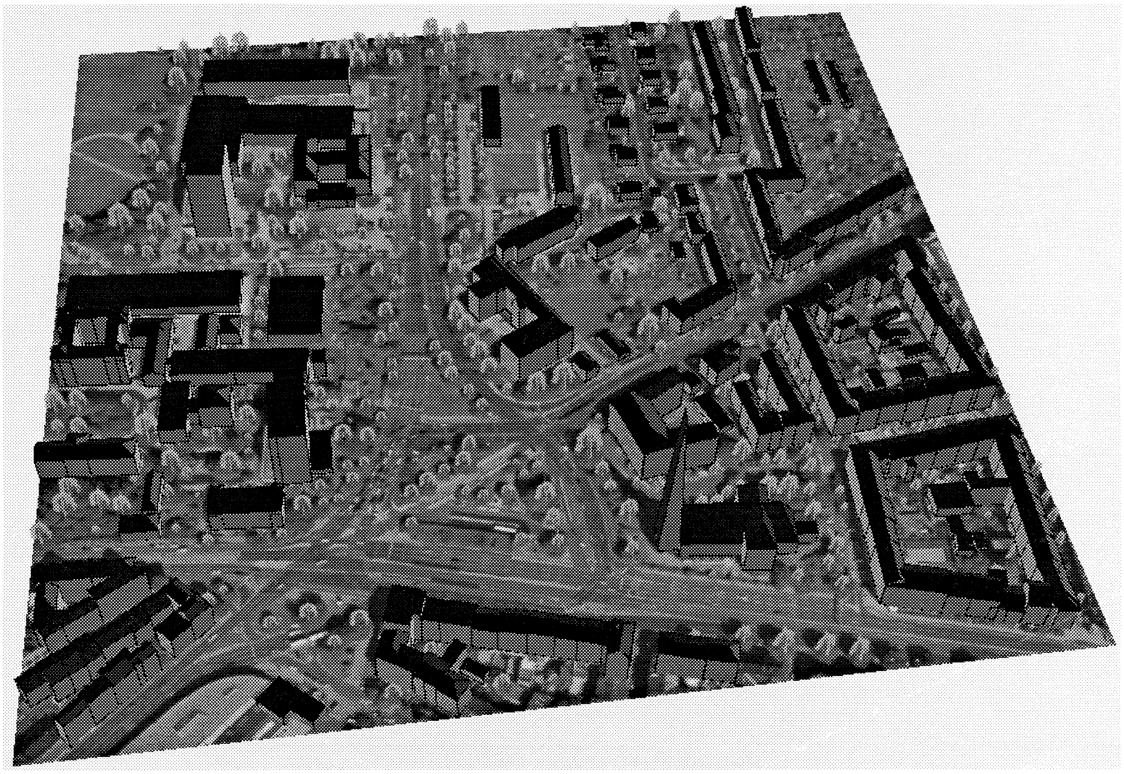

Fig. 6. 3D visualization of virtual trees and reconstructed buildings.

sion of each cuboid are already defined by the parameters of the rectangle. Remaining unknown parameters are the height of the cuboid, the roof type and roof plane slopes. These parameters are estimated by a least squares adjustment, which minimizes the distances between the DSM surface and the corresponding building primitive, i.e., the building primitives are fit to the DSM surface. The algo-

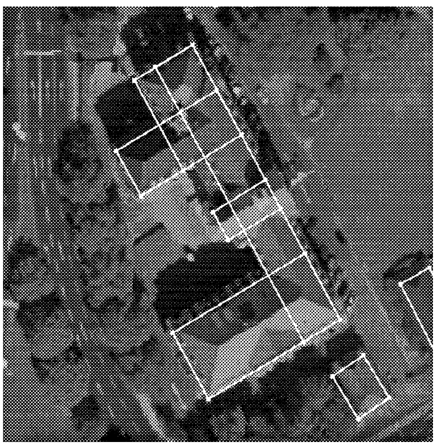

Fig. 7. Ground plan decomposed into rectangular parts.

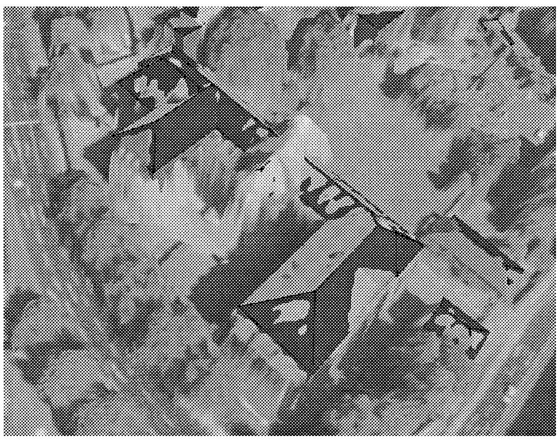

Fig. 8. Reconstructed building and DSM.

Fig. 9. 3D visualization of virtual city model.

rithm is described in more detail in (Haala et al., 1998). Fig. 8 shows the reconstructed building and the height data obtained from laser scanning. For visualization, the DSM is overlaid with the orthoimage.

Usually, simulations or visualizations require boundary representations of the buildings, which describe spatial objects by their bounding elements, e.g., planar faces. Nodes and edges are defined by intersection of the bounding planes; the topology is additionally captured by a set of relations that indicate how the faces, edges and vertices are connected to each other. In order to transform the reconstructed CSG description into a boundary representation the union of the set of CSG primitives has to be computed. For this purpose, the primitives are intersected, coplanar and touching faces are merged and inner faces or parts are removed. The result of the automatic building reconstruction for the test area is presented in Fig. 6. For the refinement of the automatic reconstruction a graphical user interface has been developed (Brenner and Haala, 1998). Using this tool, rectangles usually generated by the ground plan decomposition can, e.g., be modified or added interactively, if parts of the building like a bay are not represented in the original ground plan. The rectangles again trigger the reconstruction of an additional or modified building primitive by applying the described estimation process. In case no ground plans are available this interactive mode can also be used for semi-automatic reconstruction.

The final step towards a photorealistic visualization is presented in Fig. 9. For the generation of a virtual reality model, terrestrial images were mapped on the building faces. In this application, texture mapping from terrestrial images is simplified considerably compared to the standard architectural photogrammetry approach, since the vertices of the already reconstructed buildings provide sufficient control point information.

4. Conclusions

Data collection in urban environments can profit considerably if laser scanning is applied. The separation of spectral data into regions representing buildings, trees and grass-covered areas can be improved significantly, if a so-called normalized DSM is utilized as an additional channel within classification. This normalized DSM gives information on the height above ground for each image pixel and can be derived from laser scanning data. In the second example, the use of height data for 3D building reconstruction has been demonstrated. By the integration of ground plans, detailed reconstructions of buildings can be obtained automatically even for laser data measured at relatively low point densities. Since the integration of laser data proved to be very successful, the use of this information is strongly recommended during the automatic generation of urban databases.

References

Brenner, C., Haala, N., 1998. Fast production of virtual reality city models. Vol. 32, IAPRS, Part 4, Stuttgart, pp. 77–84.

Brunn, A., Weidner, U., 1997. Extracting buildings from digital surface models. Vol. 32, IAPRS, Part 3-4W2, Stuttgart, Germany, pp. 27–34.

Englert, R., Gülch, E., 1996. One-eye stereo system for the acquisition of complex 3D building descriptions. GEO-INFORMATION-SYSTEMS 9 (4), 16–21.

Ford, S., Kalp, D., McGlone, C., McKeown, D., 1997. Preliminary results on the analysis of HYDICE data for information fusion in cartographic feature extraction. Vol. 3072, SPIE Conference on Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III, Orlando, FL, pp. 67–86.

Haala, N., 1994. Detection of buildings by fusion of range and image data. Vol. 30, IAPRS, Part 3/1, München, pp. 341–346.

Haala, N., Brenner, C., Anders, K.-H., 1998. 3D Urban GIS from Laser Altimeter and 2D Map Data. Vol. 32, IAPRS, Part 3/1, Columbus, OH, pp. 339–346.

Henricson, O., Bignone, F., Willhuhn, W., Ade, F., Kuebler, O., Baltsavias, E., Mason, S., Gruen, A., 1996. Project Amobe: Strategies, current status and future work. Vol. 31, IAPRS, Part B3, Wien, pp. 321–330.

Hug, C., 1997. Extracting artificial surface objects from airborne laser scanner data. In: Gruen, A., Baltsavias, E., Henricson, O. (Eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images (II). Birkhäuser Verlag, Basel, pp. 203–212.

Richards, J.A., 1993. Remote Sensing and Digital Image Analysis. Springer-Verlag, Berlin.

Weidner, U., Förstner, W., 1995. Towards automatic building extraction from high resolution digital elevation models. ISPRS J. 50 (4), 38–49.