EXTRACTION OF BUILDINGS AND TREES USING IMAGES AND LIDAR DATA

N. Demir[*], D. Poli, E. Baltsavias

ETH Zurich, Institute of Geodesy and Photogrammetry, Wolfgang Pauli Str. 15, CH-8093 Zurich,
Switzerland-(demir, poli, manos)@geod.baug.ethz.ch

Commission IV, WG IV/3

KEY WORDS: DTMs/DSMs, Lidar Data Processing, Multispectral Classification, Image Matching, Information Fusion, Object Extraction, Buildings, Trees

ABSTRACT:

The automatic detection and 3D modeling of objects at airports is an important issue for the EU FP6 project PEGASE. PEGASE is a feasibility study of a new 24/7 navigation system, which could either replace or complement existing systems and would allow a three-dimensional truly autonomous aircraft landing and take-off primarily for airplanes and secondary for helicopters. In this work, we focus on the extraction of man-made structures, especially buildings, by combining information from aerial images and Lidar data. We applied four different methods on a dataset located at Zurich Airport, Switzerland. The first method is based on DSM/DTM comparison in combination with NDVI analysis; the second one is a supervised multispectral classification refined with height information from Lidar data. The third approach uses voids in Lidar DTM and NDVI classification, while the last method is based on the analysis of the vertical density of the raw Lidar DSM data. The accuracy of the building extraction process was evaluated by comparing the results with reference data and computing the percentage of data correctly extracted and the percentage of missing reference data. The results are reported and commented.

1. INTRODUCTION

Pegase is a EU FP6 project aiming at studying the feasibility of a new 24/7 navigation system, which could either replace or complement existing systems and would allow a three-dimensional truly autonomous aircraft landing and take-off primarily for airplanes and secondary for helicopters (Pegase Web, 2008). Within the project, the acquisition of a reliable geospatial reference database of the airports, and in particular the automatic extraction of buildings and obstacles at airports, has a critical role for aviation safety. Often, 3D information of airports is not available, is not accurate enough, not complete, or not updated. Thus, methods are needed for generation of accurate and complete 3D geodata with high degree of automation. Buildings and trees are considered as obstacles, so they should be correctly detected. There are several methods applied for this purpose, based on image and/or airborne Lidar data. In our approach, objects are detected and extracted in aerial images and Lidar data through a combination of multispectral image classification, DSMs and DTMs comparisons and density analysis of the raw Lidar point cloud. This paper will give a brief overview of the related work on this subject. Then, after the description of the test area at Zurich Airport, Switzerland, the strategy and algorithms will be presented and the results will be reported, compared and commented.

2. LITERATURE REVIEW

Aerial images and Lidar data are common sources for object extraction. Regarding image-based analysis, multispectral classification uses spectral properties of objects and tries to extract them by using supervised or unsupervised methods. Currently, there are commercial systems (e.g. ENVI, ERDAS, e-Cognition, IDRISI, etc.), which use expert classification algorithms such as rule-based, object-oriented and fuzzy

approaches. In digital photogrammetry, features of objects are extracted using 3D information from image matching or DSM/DTM data, spectral, textural and other information sources. Henricsson and Baltsavias (1997) combine image matching and spectral properties for the extraction process. Rottensteiner (2001) uses a semi-automated approach based on feature-based matching techniques. In Weidner and Foerstner (1995), above-ground objects are estimated using a normalised DSM (nDSM) by subtracting a DTM, estimated by a morphological filtering of the DSM, from the original DSM. In general, the major difficulty using aerial images is the complexity and variability of objects and their form, especially in suburban and densely populated urban regions.

Regarding Lidar, building and tree extraction is basically a filtering problem in the DSM (raw or interpolated) data. Some algorithms use raw data (Sohn and Dowman, 2002; Roggero, 2001; Axelsson, 2001; Vosselman and Maas, 2001; Sithole, 2001; Pfeifer et al., 1998), while some others use interpolated data (Elmqvist et al., 2001; Brovelli et al., 2002; Wack and Wimmer, 2002). The use of raw or interpolated data can influence the performance of the filtering, but also its speed being slower for raw data. The algorithms differ also in the number of points they use at a time. In addition, every filter makes an assumption about the structure of bare-earth points in a local neighbourhood. This assumption forms the concept of the filter. Often, segmentation is performed to find the clusters which delineate objects and not facets of objects. The mostly used segmentation methods are based on region-based methods, like in Brovelli et al. (2002), Crosilla et al. (2005), or use Hough transform (Tarsha-Kurdi et al., 2007). In Elmqvist et al. (2001), the terrain surface is estimated by using active contours. The performance of the algorithm is good for large objects and all types of vegetation but small objects can not be always extracted (Sithole and Vosselman, 2003). Sohn and Dowman

---

[*] Corresponding author.

(2002) separate ground and above-ground objects by a method that recursively fragments the entire LIDAR data into a set of piecewise planar surface models in order to make underlying terrain slope variations regularized into homogeneous plane terrain. Its performance is good for large buildings and all types of vegetation but poor for buildings located on slopes and low vegetation (Sithole and Vosselman, 2003). Roggero (2002) first separates ground from above-ground objects, and then from the latter dataset, buildings and trees are extracted. The method has been tested at various test sites but it is very sensitive to the input parameter values. Brovelli et al. (2002) applies edge detection before region growing, and then the terrain model is computed. Buildings and trees are extracted after a region growing process. Wack and Wimmer (2002) interpolate the original data to a regular grid and use a hierarchical approach in combination with a weighting function for the detection of raster elements that contain no ground data. The weighting function considers the terrain shape, as well as the distribution of the data points within a raster element.

Many researchers use 2D maps as prior information for building extraction in Lidar data. In Haala and Brenner (1997), geometric primitives were estimated based on histogram analysis. In general, in order to overcome the limitations of image-based and Lidar-based techniques, it is of advantage to use a combination of these techniques (Ackermann, 1999). Building reconstruction fusing Lidar data and aerial images was presented in Rottensteiner and Briese (2003). Firstly, they detected building regions in raw data, then, roofs were detected using a curvature-based segmentation technique. Additional planar faces were estimated with aerial images. Sohn and Dowman (2007) used IKONOS images to find building regions before extracting them from Lidar data. Straub (2004) combines information from infrared imagery and Lidar data to extract trees.

Few commercial software allow automatic object extraction from Lidar data. In TerraSCAN, a TIN is generated and progressively densified, using as parameters the angle points to make the TIN facets and distance to nearby facet nodes (Axelsson, 2001). In SCOP++, robust methods operate on the original data points and allow the simultaneous elimination of off-terrain points and terrain surface modelling (Kraus and Pfeifer, 1998).

In summary, most approaches try to find objects using single methods. In our strategy, we use different methods using all data with the main aim to improve the results of one method by exploiting the results from the remaining ones.

3. INPUT DATA AND PREPROCESSING

The methods presented in this paper have been tested on a dataset located in the Zurich airport. The area has an extension of 1.6km x 1.2km (Figure 1). The available data for this region are:
3D vector data of airport objects (buildings)
Colour and CIR (Colour InfraRed) images
Lidar DSM/DTM data (raw and grid interpolated)

3.1 Vector data

The 3D vector data describe buildings with 20 cm vertical accuracy. It has been produced from stereo aerial images

(Section 3.2) using a semi-automatic approach using the CC-Modeler software (Gruen and Wang, 1998, 2001).

Figure 1. Aerial image of Zurich Airport.

3.2 Image data

RGB and CIR images were acquired with the characteristics given in Table 1.

	RGB	CIR
Camera Type	Analogue	
Focal Length	303 mm	
Scale	1:10150	1: 6000
Forward overlap	70%	70%
Side overlap	26%	26%
Scan resolution	20 micron	
GSD (calculated)	12.50cm	7.25cm
Date of acquisition	July 2002	July 2002

Table 1. Main characteristics of RGB and CIR flights.

The images have been first radiometrically preprocessed (noise reduction and contrast enhancement), then the DSM was generated with the software package SAT-PP, developed at the Institute of Geodesy and Photogrammetry, ETH Zurich (Gruen and Zhang, 2003; Zhang, 2005). The matching uses multiple matching primitives and multi-image matching methods and automatic quality control. A detailed description of the matching method is given in Zhang (2005). For the selection of the optimum band for matching, we considered the image ground resolution, and the quality of each channel based on visual checking and histogram statistics. Finally, the NIR band was selected for DSM generation. The final DSM was generated with 50cm grid spacing.

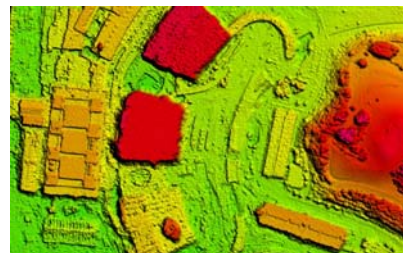Figure 2 shows a part of the generated DSM.

Figure 2. Matching DSM at Zurich Airport.

Using this DSM, CIR orthoimages were produced with 12.5cm ground sampling distance.

3.3 Lidar data

Lidar raw data (DTM-AV and DSM-AV) have been acquired with "leaves off" in February 2002 by Swisstopo. The DSM-AV point cloud includes all Lidar points (including points on terrain, tree branches etc.) and has an average point density of 1 point per 2 m^2 (Figure 3). The DTM-AV data includes only points on the ground, so it has holes at building positions and less density at tree positions (Figure 4). The height accuracy (one standard deviation) is 0.5 m generally, and 1.5 m at trees and buildings. The 2m spacing grid DSM and DTM were generated by Swisstopo with the Terrascan commercial software from the original raw data.
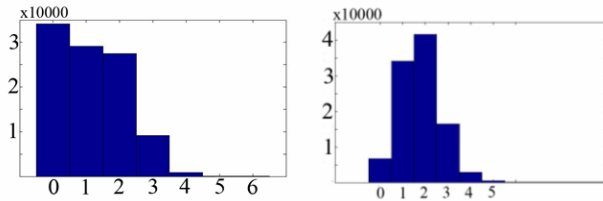

Figure 3. Density histogram (points per 2m^2) of DTM raw data (left) and DSM raw data (right) of Zurich airport. DTM and DSM have a point density of 1 and 1.5-2 points per 2m^2 respectively.

Figure 4. DTM-AV raw data of Zurich airport. The white areas represent voids (buildings, aircrafts, vehicles, high or dense trees).

4. BUILDING EXTRACTION

Four different approaches have been applied to exploit the information contained in the image and Lidar data, extract different classes of objects and finally buildings. The first method is based on DSM/DTM comparison in combination with NDVI analysis for building extraction. The second approach is a supervised multispectral classification refined with height information from Lidar data. The third method uses voids in Lidar DTM and NDVI classification. The last method is based on the analysis of the vertical density of the raw DSM Lidar data. The accuracy of the building extraction process was evaluated by comparing the results with the reference data and computing the percentage of data correctly extracted and the percentage of reference data not extracted.

4.1 DSM/DTM and NDVI

By subtracting the DTM from the DSM, a so-called normalized DSM (nDSM) is generated, which describes the above-ground objects, including buildings and trees (Zhang, 2003). As DSM, the surface model generated by SAT-PP and as DTM the Lidar DTM grid were used. Some details of the nDSM are shown in Figure 5. As it can be seen, the nDSM is quite good including also small trees, bushes and vehicles .
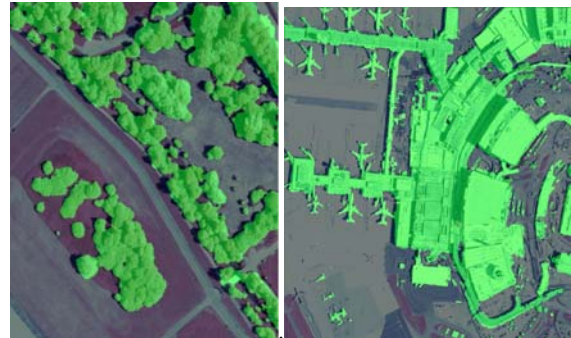

Figure 5. nDSM overlaid on orthoimage.

A standard unsupervised (ISODATA) classification of the CIR orthoimage was used to compute an NDVI image, containing vegetation (mainly trees and grass). The intersection of the nDSM with NDVI corresponds to trees. By subtracting the resulting trees from the nDSM, the buildings are obtained. 76% of building class pixels were correctly classified, while 10% of the reference data were not detected. The main reason for the 76% is green material on buildings, espec. roofs. The extracted buildings are shown in Figure 6. In the final result, some non-building objects are remaining such as aircrafts and vehicles.

Figure 6. Building class (in red) extracted from intersection of nDSM and NDVI.

4.2 Supervised classification

The basic idea of this method is to combine the results from a supervised classification with the height information contained in the nDSM. Supervised classification methods are preferable to unsupervised ones, because the target of the project is to detect well-defined standard target classes (airport buildings, airport corridors, bare ground, grass, trees, roads, residential houses, shadows), present at airport sites. The training areas were selected manually using AOI tools within the ERDAS Imagine commercial software (Kloer, 1994).

Among the available image bands for classification (R, G and B from colour images and NIR, R and G bands from CIR images), only the bands from CIR images were used due to their better resolution and the presence of NIR channel (indispensable for grass and trees extraction). In addition, new synthetic bands were generated from the selected channels: a) 3 images from principal component analysis (PC1, PC2, PC3); b) two images from NDVI computation using the NIR-R and R-G channels (NDVI1 and NDVI2 respectively) and c) one saturation image (called S) obtained by converting the NIR-R-G channels in the IHS (Intensity, Hue, Saturation) colour space.

The separability of the target classes was analyzed through use of plots of mean and standard deviation for each class and channel and divergence matrix analysis of all possible combinations of the three CIR channels and the additional channels, mentioned above. The analysis showed that:

- G and PC2 have high correlation with other bands
- NIR-R-PC1 is the best combination based on the plot analysis
- NIR band shows good separability based on the divergence analysis
- PC1-NDVI-S combination shows best separability over three-band combinations based on the divergence analysis

Therefore, the combination NIR-R-PC1-NDVI-S was selected for classification. The maximum likelihood classification method was used. As expected from their low values in the divergence matrix, grass and trees, airport buildings and residential houses, airport corridors and bare ground, airport buildings and bare ground could not be separated. Using the height information from nDSM, airport ground and bare ground were fused into "ground" and airport buildings with residential houses into "buildings", while trees and grass, as well as buildings and ground could be separated. The final classification is shown in
Figure 7. 86% of building class is correctly classified, while 13% of the reference building data were not detected. Aircrafts and vehicles are again mixed with buildings.
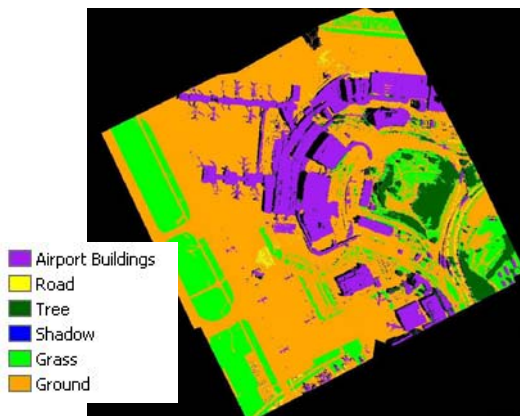

Figure 7. Results from supervised classification combined with height information.

4.3 Building extraction using density of DTM-AV and NDVI

Buildings and other objects, like high or dense trees, vehicles, aircrafts, etc. are characterized by null or very low density in the DTM-AV point cloud. Using the vegetation class from NDVI channel as a mask, the areas covered by trees are

eliminated, while small objects (aircrafts, vehicles) are eliminated by deleting them, if their area was smaller than $25m^2$. Figure 8 shows the final building class. 87% of building class pixels are correctly classified, while 13% of the reference data are not detected.

Figure 8. Lidar DSM points located on buildings, extracted using DTM-AV voids and NDVI.

4.4 Building and tree extraction from Lidar data

As mentioned above, in DSM raw data the point density is generally much higher at trees than at open terrain or buildings. On the other hand, tree areas have low density in DTM-AV data, as it can be seen in Figure 4. We start from regions that are voids or have low density in the raw DTM. These regions represent mainly buildings and trees and are used as mask to select the raw DSM points for further analysis.

The first step is identification of buildings by fitting planes, and the elimination of these points. The reason is that building roofs may cause problems in the subsequent step of vertical point density and distribution analysis, aimed at identification of trees. The plane fit operation is possible with different commercial software. Here, Geomagic Studio by (Geomagic Studio, Raindrop Geomagic, Research Triangle Park, NC) was used. The planes of small buildings and non-planar parts of large buildings could not be detected. With the remaining points, the analysis regarding vertical point density and distribution is applied in search windows with size 2.5 m x 2.5 m. This size was selected larger than the average point density of the DSM. From Figure 3, we see that the average density is 1.5-2 points for a window of 1.4 m x 1.4 m.

The points in each search window are projected onto the xz plane and divided in four equal subregions, using x_{min}, x_{max}, z_{min} and z_{max} as boundary values, with $x_{max} = x_{min} + 2.5m$. The density in the four subregions is computed (see example in Figure 9).
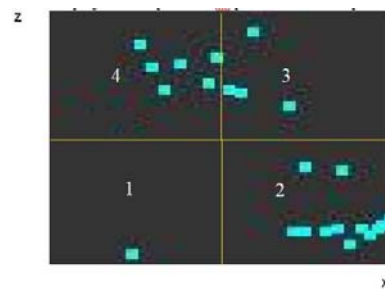

Figure 9. Projected points in xz plane and 4 subregions.

Point clouds in each window are automatically classified as trees, for certain combinations of point density in the four subregions. These combinations generally request that at least one upper subregion has high density, while at least one lower subregion should have lower density. In addition, we request that the difference Zmax – Zmin in each window exceeds a threshold. For the upper subregions a threshold *T* for the point density is used, where *T* is defined as:

$$T = \mu_{UP} + \frac{\sigma_{UP}}{2} \qquad (1)$$

$_{UP}$ and $_{UP}$ are the mean value and standard deviation of the number of points in each of the upper subregions in all the windows.
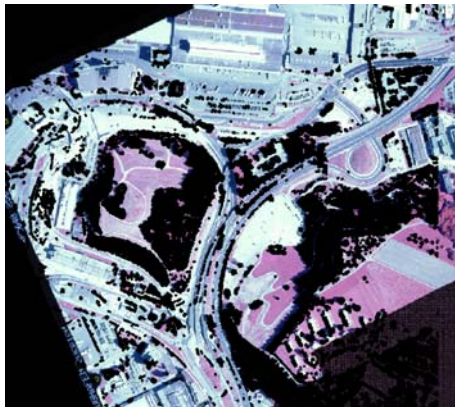

Figure 10. Extracted trees (black points) overlaid on CIR orthoimage.

In the case of Zurich Airport data, the density threshold *T* is 6 points / 2.5m by 2.5m. The extracted tree class has been compared with the tree class extracted from NDVI and nDSM. 73% of tree points were correctly classified, while 8% were not detected. The density of point cloud directly affects the quality of the result. As it can be seen in Figure 10, visible errors in the results are small objects as vehicles and aircrafts. In addition, some tree areas could not be extracted because of the low point density in the whole dataset. As mentioned above, the points in the raw DSM not present in DTM-AV describe buildings, vehicles and high or dense vegetation. After extracting the trees using point density analysis, buildings are obtained by subtraction of the tree layer from the DSM points, corresponding to voids and low density, and filtering of small objects (Figure 11). The accuracy analysis shows that 92% of building pixels are correctly classified, while 17% of buildings could not be detected.
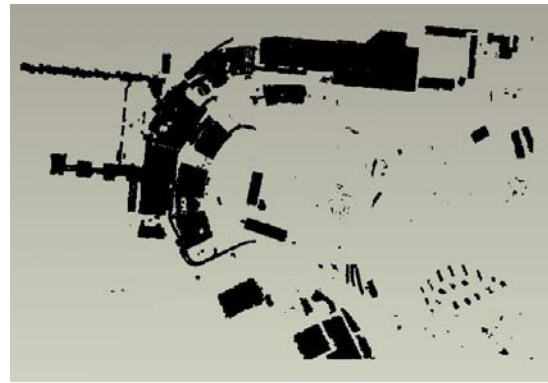

Figure 11. Extracted building points after elimination of tree points.

5. ANALYSIS OF RESULTS

The accuracy results of the four methods described in Section 4 are summarized in Table 2. Method 4, based only on Lidar data, performs best in terms of correctness, but is the worst in terms of completeness. It does not detect all buildings, but those detected are correct. On the other hand, Method 1, again based on Lidar data, but with NDVI contribution, can extract the largest part of buildings but other objects are included, resulting in the worst correctness value. The other two methods have basically the same performance, lying in the middle between Method 1 and 4 results. It should be noted that results using Lidar data strongly depend on average point density, but also number of echoes that are registered per pulse and whether Lidar data acquisition was with leaves on or off.

Due to time restrictions, only a first simple fusion of the results has been attempted. By union of the four building detection results, the omission rate decreases (8%) but also correctness too (81%), while the intersection of all results gives the best correctness (96%). The correctness of each method could be improved by developing an automatic detection of objects like aircrafts, which are classified as buildings in all methods. Taking into account the advantages and limitations of each method, currently we can not recommend a single solution for building extraction. By intersecting the results from method 4, based on Lidar data analysis, and method 2, based on supervised classification, the best correctness rate is achieved, but the completeness is poor. The other building layer combinations in Table 2 led to worse results.

	Method 1	Method 2	Method 3	Method 4	1∪2∪3∪4	1∩2∩3∩4	1∪4	1∩4	2∪4	2∩4
	nDSM+NDVI	Class.+nDSM	Voids+NDVI	Lidar						
Correctness (%)	76	86	87	**92**	81	**96**	83	96	83	**95**
Omission error (%)	**10**	13	13	17	**8**	29	10	29	**9**	27

Table 2. Results of building extraction using four methods and combinations thereof. The best results are shown in bold.

6. CONCLUSIONS

In this paper, different methods for object extraction (mainly buildings) in Lidar data and aerial images have been presented. In each method, the basic idea was to get first preliminary results and improve them later using the other available data. The methods have been tested on a dataset located at Zurich Airport, Switzerland, containing aerial colour and infrared images, Lidar DTM and DSM point clouds and regular grids and vector data for accuracy assessment. The results showed

that correctness values up to 92% can be achieved using Lidar data only, while the highest completeness is obtained by the combination of image and Lidar data.

Future work will include the improvement of building extraction from aerial images and Lidar data. The algorithms will be tested also on other airport locations. However, the main focus will be on a better fusion of the individual results, use of image edges for better building delineation and more detailed building modelling.

REFERENCES

Ackermann, F., 1999. Airborne Laser Scanning - Present Status and Future Expectations. ISPRS Journal of Photogrammetry and Remote Sensing, 54 (2-3), pp. 64-67.

Axelsson, P., 2001. Ground estimation of laser data using adaptive TIN-models. Proc. of OEEPE workshop on airborne laserscanning and interferometric SAR for detailed digital elevation models, pp. 185-208.

Brovelli, M.A., Cannata, M., Longoni U.M., 2002. Managing and processing Lidar data within GRASS. Proc. of the GRASS Users Conference 2002, Trento, Italy.

Crosilla, G., Visintini, D., Sepic, F., 2005. A Segmentation procedure of Lidar data by applying mixed parametric and nonparametric models. IAPRS[*], Vol. 36, Part 3/W19.

Elmqvist, M., Jungert, E., Lantz, F., Persson, A., Soderman, U., 2001. Terrain modelling and analysis using laser scanner data. IAPRS[*], Vol. 34, Part 3/W4, pp. 219-227.

Gruen, A., Wang, X., 1998. CC-Modeler: A topology generator for 3-D city models. ISPRS Journal of Photogrammetry & Remote Sensing, 53(5), pp. 286-295.
Gruen, A., Wang, X. 2001. News from CyberCity-Modeler. In: Automatic Extraction of man-made objects from aerial and space images (III), E. Baltsavias, A. Gruen, L. Van Gool (Eds.), Balkema, Lisse.

Gruen, A., Zhang, L., 2003. Sensor Modeling for Aerial Triangulation with Three-Line-Scanner (TLS) Imagery. Photogrammetrie, Fernerkundung, Geoinformation, (2), pp. 85-98.

Haala, N., Brenner, C., 1997. Generation of 3D city models from airborne laser scanning data. Proc. EARSEL workshop on Lidar remote sensing on land and sea, Tallinn, Estonia, 8 p.

Henricsson, O., Baltsavias, E., 1997. 3D Building reconstruction with ARUBA: a qualitative and quantitative evaluation. In: Proc. of the international Workshop "Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)", 5. -9. May, Ascona, Switzerland, A. Gruen, E. P. Baltsavias and O. Henricsson (Eds.), Birkhäuser Verlag, Basel, pp. 65-76.

Kloer, B. R., 1994. Hybrid Parametric/Non-Parametric Image Classification. Paper presented at the ACSM-ASPRS Annual Convention, Reno, Nevada, April.

Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing, 53, pp. 193-203.

Pegase Web, 2008. http://dassault.ddo.net/pegase (last visited May 2008).

Pfeifer, N., Kostli, A., Kraus, K., 1998. Interpolation and filtering of laser scanner data - implementation and first results. IAPRS[*], Vol. 32, Part 3/1, pp.153 - 159.

Roggero, M., 2001. Airborne Laser Scanning: Clustering in raw data. IAPRS[*], Vol. 34, Part 3/W4, pp. 227-232.

Rottensteiner, F., 2001. Semi-automatic extraction of buildings based on hybrid adjustment using 3D surface models and management of building data in a TIS. PhD Thesis, Institute of Photogrammetry and Remote Sensing, Vienna University of Technology.

Rottensteiner, F., Briese, C., 2003. Automatic generation of building models from Lidar data and the integration of aerial images. IAPRS[*], Vol. 34, Part 3-W13, pp. 174-180.

Sithole, G., 2001. Filtering of laser altimetry data using a slope adaptive filter. IAPRS[*], Vol. 34, Part 3/W4, pp. 203-210.

Sithole, G., Vosselman, G., 2003. Report: ISPRS Comparison of Filters, Department of Geodesy, Faculty of Civil Engineering and Geosciences, Delft University of Technology, The Netherlands.

Sohn, G., Dowman, I., 2002. Terrain surface reconstruction by the use of tetrahedron model with the MDL criterion. IAPRS[*], Vol. 34, Part 3A, pp. 336-344.

Sohn, G., Dowman, I., 2007. Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction. ISPRS Journal of Photogrammetry and Remote Sensing, 62(1), pp. 43-6.

Straub, B.-M., 2004. A multiscale approach for the automatic extraction of tree tops from remote sensing data. IAPRS[*], Vol. 35, Part B3, pp. 418-423.

Tarsha-Kurdi, F., Landes, T., Grussenmeyer, P., 2007. Hough-transform and extended RANSAC algorithms for automatic detection of 3D building roof planes from Lidar data. IAPRS[*], Vol. 36, Part 3/W52.

Vosselman, G., Maas, H., 2001. Adjustment and Filtering of Raw Laser Altimetry Data. Proc. of OEEPE workshop on airborne laserscanning and interferometric SAR for detailed digital elevation models, 1-3 March, 11 p.

Wack R., Wimmer A., 2002. Digital Terrain Models From Airborne Laser Scanner Data – A Grid Based Approach. IAPRS[*], Vol 34, Part 3B, pp. 293-296.

Weidner, U., Foerstner, W., 1995. Towards automatic building extraction from high-resolution digital elevation models. ISPRS Journal of Photogrammetry and Remote Sensing, 50(4), pp. 38–49.

Zhang, L., 2005. Automatic Digital Surface Model (DSM) generation from linear array images. Ph.D. thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Report No. 90.

[*] IAPRS: International Archives of The Photogrammetry, Remote Sensing and Spatial Information Sciences.