14th Annual International Conference and Exhibition on Geospatial InformationTechnology and Applications

INDIA GEOSPATIAL FORUM

7-9 February, 2012

THEME: Geo-Budget: Enabling Sustainable Growth

# Extracting single trees with airborne lidar data

SANDEEP GUPTA[1]*, BARBARA KOCH[1]

[1]Department of Remote Sensing and Landscape Information Systems, Albert-Ludwigs University, Tennenbacher str. 4, 79106 Freiburg, Germany
Phone: +49-761 203 3695, 3699; Fax: +49-761 203 3701

*Corresponding author and presenter. E-mail: sandeep.gupta@felis.uni-freiburg.de

Paper Reference Number: PN-104

*Summary*

The results of single tree cluster extraction with point cloud derived from full waveform airborne laser scanning (ALS) in black forest area of Germany using k-means algorithm has been presented in this work. The validation of 3-D tree top candidates with field inventoried reference tree tops with a sum of 289 trees illustrate that the detection rate for pine and oak species is 76% and 68%, respectively. There was a 64% and 56% one-to-one linking between candidate and field tree tops of the pine and oak species, respectively. The form and density of the canopy played significant role in achieving average accuracy, beside the limitation of algorithm performance. The validation of crown height as an additional parameter could be done only for two plots of oak species, which shows an average variation of 6 m between the linked candidate and reference trees. This could be due to the lower returns of ALS points in wider and dense canopy of oak trees. The outcome shows the potential of algorithm in the single tree laser remote sensing.

## Introduction

Aerial scanning with Light Detection and Ranging (LiDAR) data has been in use since more than a decade to derive forest information at stand and/or tree level (Hyyppä *et al.*, 2001; Persson *et al.*, 2002). Depending on the requirement, airborne laser scanner (ALS) data are analyzed using two main approaches: area/normalized digital surface model (nDSM)-based and individual tree crown based approach (Maltamo *et al.*, 2006). The individual tree approach are either used to derive tree attributes such as crown height, diameter and volume, stem volume etc. or the tree level information are aggregated to derive stand level information such as forest height, timber volume, biomass. Various studies in the past has been conducted on the LiDAR based vegetation information extraction using different methods (Nilsson, 1996; Hyyppä and Inkinen, 1999; Persson *et al.*, 2002; Hyyppä *et al.*, 2006; Wang *et al.*, 2008; Vauhkonen *et al.*, 2011). Retrieving information of the trees lying below the upper canopy layer is a trivial task and is the major disadvantage associated with nDSM based approach (Reitberger *et al.,* 2008a). Reitberger *et al.* (2008a) showed an improved tree detection rate by applying a segmentation technique based on the normalized cut segmentation method on the full waveform LiDAR points of different density. However, as the detection rate increases, false detections increase as well, doubting thus on the approach reliability (Pitkänen *et al.,* 2004; Reitberger *et al.,* 2008a, Vauhkonen *et al.,* 2011). Several tree top detection studies with some adaptations using local maxima have been carried out in the past and majority of them are raster based (Persson *et al.,* 2002; Popescu *et al.,* 2002; Pitkänen *et al.,* 2004; Koch *et al.,* 2006; Tiede *et al.,* 2008).

Among the vector-based methods, clustering is one of the promising approaches for single tree detection (Gupta *et al.* 2010). The *k*-means is one of the most popular iterative partitioning based clustering approaches. In this, vector data of an area of interest are being partitioned into a group of clusters using a distance criterion (Jain *et al.,* 1999). Single tree laser remote sensing with various clustering mechanisms has been reported in the past (Morsdorf *et al.,* 2003; Morsdorf *et al.,* 2004; Cici *et al.,* 2008; Doo-Ahn *et al.,* 2008; Reitberger *et al.,* 2008b). Morsdorf *et al.* (2003) conducted clustering based study in the mountain Pine dominated boreal forest of Swiss National Park. In their study, they used first and last pulse ALS data of high density (30 points $m^{-2}$), local maxima derived from digital surface model (DSM) as seed points and supplied the two in to *k*-means algorithm for single trees extraction. In contrast to the approach of the presented work, instead of scaling-down the z-coordinates, Morsdorf *et al.* (2003) scaled it up by a factor of three to accommodate the aspect ratio (height to width ratio) of pine tree crowns, which ranged from 3-6. It is noteworthy that the *k*-means method works well when a data set has “compact” or “isolated” clusters (Mao and Jain, 1996). Hence, it is preferential to downsize the height value of the normalized LiDAR points and the corresponding seed points to minimize the intra-cluster variance, which is the eventual objective of the *k*-means approach. Additionally, the presented study differs from Morsdorf *et al.* (2003) in three ways: first, the study focus on the both Pine and Oak dominated and mixed species plots in temperate forest with multilayered structure; second, the superfluous local maxima points obtained in the step before main processing were removed by applying a search algorithm and third, the study also compare the crown height, an additional parameter, for two plots of Oak.

The objective of the work presented in this study is to assess the applicability of a supervised *k*-means approach in single tree detection using ALS data in both dominant and mixed area containing Pine (coniferous) and Oak (deciduous) species of different maturity situated in the temperate black forest region of Germany. The accuracy was assessed using a well defined validation procedure.

### *Study area*

The chosen test area is situated in the black forest area near Karlsruhe, Germany. The coordinates in Gauss Krüger system and in meters for the upper left corner are 3456375, 5433820 and the lower right corner 3458025, 5432980. The area include a total of 289 trees containing Pine *(Pinus sylvestris)* and Oak species (*Quercus rubra* and *Quercus petraea*) of different maturity level with high canopy density.

## Materials and Methodology

### *Field measurements*

The Forest Research Institute (FVA) of the Federal State of Baden-Württemberg, Germany, offered to use the Forest inventory data collected during summer 2006. Another field inventory was carried during November 2010 to collect tree level information for additional areas. In the field, all the trees above 7 cm of breast height diameter (DBH) were measured.

*LiDAR data characteristics*

Two sets of ALS data were used in this study. The first set of full waveform derived ALS data was recorded in August 2007 by TopoSys GmbH using a "Harrier 56/G4" Riegl LMS-Q56O scanner mounted on a helicopter and flown at a height of 450 m above ground level. To achieve a high point density (16 points $m^{-2}$), the study area was scanned twice during day and night with more than 50% of overlapping (Straub *et al.,* 2009). The second set of IGI full waveform derived ALS data of 20 points $m^{-2}$ was collected during November 2009. The digital models (DTM, DSM and nDSM) were calculated in TreesVis – a LiDAR visualization, processing and analysis software developed in FeLis (Weinacker *et al.,* 2004). Normalization over the raw laser scanner data were performed using the DTM. It helped in obtaining the absolute height of each LiDAR point and removing the terrain-induced effects. The normalized data-set was used for clustering.

*Single tree detection*

The processing of 2007 LiDAR data and that of 2009 data was carried out in seven and five rectangular plots, each of size 50 by 50 m, respectively.

Local maxima above 5 m height were calculated using the nDSM having a gray value larger than the gray value of all its eight neighboring pixels. The superfluous local maxima detected from the multiple peaks from the crown were excluded using a search algorithm implying a threshold distance criteria. The distance between the two points is computed in Euclidean 3-D space. If the distance between the two points is less than the threshold distance defined, then the point selected is removed. This process is continued until all the local maxima points below the threshold distance are removed. The sieved local maxima were eventually used as external seed points in the *k*-means algorithm. The distance criteria are related to tree type distribution. The threshold distance for conifers with single and narrow crown at the top was relatively lower (2-4 m) than that of broadleaved trees with wider crown and more intermittent peaks at the top (4-6 m). The distance was a result of a visual analysis of the study area through the aerial photograph. The empirical study on the role of distance function is advantageous in case of large area estimation where there is a gain of pre-knowledge about the distribution of tree types.

The simple *k*-means algorithm described by Jain *et al.* (1999) was supervised in this study by using local maxima as external seed points to initialize the process. Single tree clusters containing 3-D points were obtained after performing the supervised *k*-means algorithm over normalized LiDAR data and the corresponding external seed points above the height of 5 m. The clustering above 5 m was performed to avoid the effect of low ground vegetation and other objects while processing. The supervised *k*-means has two major advantage over traditional one: first, random seed selection procedure with a trial and error based approach containing several repetitions to select an appropriate *k* clusters can be completely avoided and second, reduction in the time and machine cost. It was found that by compressing the ALS point cloud along z-axis during supervised *k*-means run-time, the point distribution gets more compact in the direction of height. This causes an improved data partitioning. The compression value was kept half, which was empirically found (Gupta *et al.,* 2010). The reason behind selecting this value was based on several result analyses criteria. It was concluded after the analysis that there is no substantial improvement in the quality of partitioned tree clusters when the compression is more than half.

*Performance Evaluation*

The performance of the algorithm was evaluated by

1. Comparing detection rate. This is done by finding the total detected trees in proportion to the reference trees
2. Calculating the “user’s accuracy” and the “producer’s accuracy” (Congalton, 1991) based on one-to-one spatial linkage up to 5 m in three dimensional Euclidean distance (3-D ED) between detected and reference tree tops
3. Estimating omission error, i.e. field tree tops that could not be linked to any of the detected tree tops and commission error, i.e. detected tree tops that could not be linked to any of the field-inventoried trees

In addition, crown height estimates for two plots dominated by Oak species were also compared. For other plots, the comparison could not be made due to non-availability of tree parameters in field data.

## Results and Discussion

Single tree clusters containing 3-D points, thus, obtained after performing the supervised *k*-means algorithm. From 3-D points of each tree cluster, the tree tops and crown height were computed. Maximum height and the corresponding position of each tree cluster are regarded as the detected 3-D tree top point and this has the maximum height among all LiDAR points belonging to the tree cluster. Accuracies of the result are presented in Table 1.

Table 1. Overall tree detection accuracies. Where, Reference = total no. of reference trees, Detection = total no. of detected tree clusters, detection rate = % of detection to reference trees, Correct detect = one-to-one linked detected tree tops with corresponding reference tree tops, User accy = user's accuracy = % of Correct detect to Detection, Prod accy = producer's accuracy = % of Correct detect to Reference, Com err = commission error or false detection = 100-User accy and Om err = omission error = 100-Prod accy.

| Year/ Plot | Reference | Detection | Detection rate | Correct detect | Prod accy | User accy | Com err | Om err |
|---|---|---|---|---|---|---|---|---|
| **2009/ 5plots** | 165 | 116 | 70.3 | 101 | 61.2 | 87.1 | 12.9 | 38.8 |
| **2007/ 7plots** | 124 | 94 | 75.8 | 74 | 59.7 | 78.7 | 21.3 | 40.3 |
| **Sum/ Avg.** | 289 | 210 | 72.7 | 175 | 60.6 | 83.3 | 16.7 | 39.4 |

It is evident from Table 1 that the average tree detection rate in the two data is approximately 70% and 76%. The higher detection in 2007 data has also influenced the increase in commission error. The 2007 result are of Pine and Oak trees selected from seven plots of mixed species and have therefore lower accuracy. The producer’s accuracies and omission errors are nearly same for all the aggregated plots of the two data sets. However, there is a substantial variation (8.4%) in user's accuracies from the result of the two data-sets obtained from pure and mixed species plots.

Table 2. Species-wise accuracies. Pine and Oak species from mixed species plots of the 2007 data and species dominant plots of 2009.

Table 2 depicts results of tree detection rate, accuracies and error for the Pine and Oak species, the two main species found in the area. As we can see, there is a great variation in the detection rate between the two species from the clusters of two different distribution and period, one obtained in August 2007 (leaf-on) from mixed species plots and another during November 2009 (nearly the end of leaf-on period) from species dominant plots. Overall, it was found that the detection rate, producer's accuracy was higher and omission error was lower for Oak dominating two plots of 2009 data compare to results from 2007 data of the same species distributed in seven plots within mixed species condition and vice-verse is true for Pine species. There were minor differences in the user's accuracy for the Oak species distributed differently in two data-sets acquired during 2007 and 2009: 81.8% and 83.3%, and that for the commission error is 18.2% and 16.7%. While, there was a substantial increase in user's accuracy for Pine species from the clusters obtained in 2007 where species are not from the dominating plots compare to results from 2009 data, i.e. 76% in mixed distribution and 89% in single species dominating plots. Reflecting the effect on commission error, there is a decrease in commission error from 24% to 11% for the Pine species in the clusters of 2007 and 2009 data. The low detection and high omission of plots dominated with Pine trees from 2009 data could be partly due to low local maxima generated which is directly related to number of tree clusters to be produced and partly due to two young plots out of the three Pine plots. Rest of the two plots of 2009 data belonged to matured Oak species. The result was however, surprising for the Pine species (conifer) but was encouraging for Oak species (broad-leaved) for its higher accuracy. The presence of crown height as an additional parameter for two 2009 field inventoried plots of matured Oak species was helpful in comparing it with detected tree clusters. It was found that there was a 6.1 m of deviation in the mean crown height of the linked detected trees with respect to reference trees. Finding the crown height in the detected tree clusters containing 3-D points is a trivial task and is depend on several factors such as canopy and tree density, branching pattern of the tree species. These factors are directly related to the number, density and intensity of the returns echoes during aerial laser scanning. In this study, a little higher deviation in the crown height estimation may be due to decrease in the number of return echoes in the original data from the lower height of the trees due to above mentioned factors. It is concluded that some more experiments are required in this direction to validate the algorithm for two species in different environment.

| **Year/Plot** | **Reference** | | **Detection** | | **Detection rate** | | **Correct detect** | | **Prod accy** | | **User accy** | | **Com err** | | **Om err** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Pine** | **Oak** | **Pine** | **Oak** | **Pine** | **Oak** | **Pine** | **Oak** | **Pine** | **Oak** | **Pine** | **Oak** | **Pine** | **Oak** | **Pine** | **Oak** |
| **2009/ 5plots** | 127 | 38 | 80 | 36 | 63.0 | 94.7 | 71 | 30 | 55.9 | 78.9 | 88.75 | 83.3 | 11.25 | 16.7 | 44.1 | 21.1 |
| **2007/ 7plots** | 44 | 80 | 50 | 44 | 113.6 | 55 | 38 | 36 | 86.4 | 45 | 76.0 | 81.8 | 24.0 | 18.2 | 13.6 | 55 |
| **Sum/ Avg.** | 171 | 118 | 130 | 80 | 76.0 | 67.8 | 109 | 66 | 63.7 | 55.9 | 83.8 | 82.5 | 16.2 | 17.5 | 36.3 | 44.1 |

## Conclusions

The supervised *k*-means is useful in partitioning of ALS point cloud in to single tree clusters. The detection of local maxima, which serve as an external seed point during the clustering process should be carefully chosen as it is directly related to commission and omission errors. By compressing the ALS point cloud along z-axis during supervised *k*-means run-time, the distribution of points gets more compact in height. This causes improved partitioned clusters. The result after validation showed that the algorithm performed quite well for Oak species compare to Pine. There was a significant increase in the detection rate, producer's accuracy and decrease in commission and omission errors for Oak species measured from two plots where it is dominantly distributed in 2009 data compare to sparsely distributed seven plots containing a mix of tree species in 2007 data. However, it should also be noted that the number of reference trees of the two species in different period also differs significantly. The number of field inventoried Pine trees during 2009 was three times more than the 2007 data, located distantly from each other; where as the number of Oak species was reduced to half. Therefore, there is a need of more field inventoried tree level information for algorithm performance evaluation. In addition to crown height, other parameter such as maximum crown diameter, crown volume and stem volume may also be considered during field inventory for better accuracy estimates.

The quality of the result depends largely on the flight parameters, the forest conditions, the LiDAR point density, the distance threshold and the number of seed points. Empirical analyses need to be tested in the process to get a best value adaptation.

Apart from improvement in the existing single tree method, future study for ALS based tree inventories will focus on the use of a combined approach where both single tree and area based approach will be used for better estimates and large area mapping. Already, the possibility to incorporate intensity value in forest stand classification of the ALS data, based on the specific rule sets, is under investigation.

*Acknowledgments*

The research was funded from EU FP7 project 'Flexwood'. The authors would like to acknowledge the state Forest Research Institute of Baden-Württemberg (FVA) for granting the access of reference data of the study area. The authors would also like to thank the anonymous reviewers for their thorough reviews.

***References***

CICI, A., KEVIN, T., NICHOLAS, J. T., SARAH, S. and JÖRG, K., 2008 Extraction of vegetation for topographic mapping from full-waveform airborne laser scanning data. In *Proceedings of SilviLaser 2008*, 17-19 September 2008, Edinburgh, U.K., pp. 343-353.

CONGALTON, R. G., 1991 Considerations and techniques for assessing the accuracy of remotely sensed data. *Remote Sensing of Environment*. 37, 35-46.

DOO-AHN, K., WOO-KYUN, L. and HYUN-KOOK, C., 2008 Estimation of effective plant area index using LiDAR data in forest of South Korea. In *Proceedings of SilviLaser 2008*, 17-19 September 2008, Edinburgh, U.K., pp. 237-246.

GUPTA, S., WEINACKER, H. and KOCH, B., 2010 Comparative analysis of clustering-based approaches for 3-d single tree detection using airborne fullwave lidar data. *Remote Sensing*. **2(4)**, 968-989.

HYYPPÄ, J. and INKINEN, M., 1999 Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*. **16**, 27-42.

HYYPPÄ, J., KELLE, O., LEHIKOINEN, M. and INKINEN, M., 2001 A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geosceince and Remote Sensing*. **39(5)**, 969-975.

HYYPPÄ, J., Yu, X., HYYPPÄ, H. and MALTAMO, M., 2006 Methods of airborne laser scanning for forest information extraction. In *EARSeL-Proceedings of International Workshop on 3D Remote Sensing in Forestry*, 14-15 February 2006, Vienna, Austria, pp. 63-78.

JAIN, A.K., MURTY, M.N. and FLYNN, P.J., 1999 Data Clustering: A Review. *ACM Computing Surveys*. **31(3)**, 264-323.

KOCH, B., HEYDER, U. and WEINACKER, H., 2006 Detection of individual tree crowns in airborne lidar data. *Photogrammetric Engineering & Remote Sensing*. **72(4)**, 357-363.

MALTAMO, M., EERIKÄINEN, K., PACKALÉN, P. and HYYPPÄ, J., 2006 Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry*. **79(2)**, 217-229.

MAO, J. and JAIN, A.K., 1996 A self-organizing network for hyperellipsoidal clustering (HEC). *IEEE Trans. Neural Netw*. **7**, 16–29.

MORSDORF, F., MEIER, E., ALLGÖWER, B. and NÜESCH, D., 2003 Clustering in airborne laser scanning raw data for segmentation of single trees. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34 (part 3/W13), Dresden, Germany, pp. 27-33.

MORSDORF, F., MEIER, E., KÖTZ, B. and ITTEN, K.I., 2004 Lidar based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment*. **92**, 353-362.

NILSSON, M., 1996 Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment*. **56(1)**, 1-7.

PERSSON, Å., HOLMGREN, J. and SÖDERMAN, U., 2002 Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering & Remote Sensing*. **68**, 925-932.

PITKÄNEN, J., MALTAMO, M., HYYPPÄ, J. and WEI YU, X., 2004 Adaptive methods for individual tree detection on airborne laser based canopy height model. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (pasrt 8/W2), Freiburg, Germany, pp. 187-191.

POPESCU, S.C., WYNNE, R.H. and NELSON, R.F., 2002 Estimating plot-level tree heights with lidar: local filtering with a canopy height based variable window size. *Computers and Electronics in Agriculture*. **37**, 71-95.

REITBERGER, J., KRZYSTEK, P. and STILLA, U., 2008a 3D segmentation and classification of single trees with full waveform LIDAR data. In *Proceedings of SilviLaser 2008*, 17-19 September 2008, Edinburgh, U.K., pp. 216-226.

REITBERGER, J., KRZYSTEK, P. and STILLA, U., 2008b Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *International Journal of Remote Sensing*. **29(5)**, 1407-1431.

STRAUB, C., DEED, M., WEINACKER, H. and KOCH, B., 2009 Using Airborne Laser Scanner Data and CIR Orthophotos to Estimate the Stem Volume of Forest Stands. *Photogrammetrie - Fernerkundung – Geoinformation*. **2009(3)**, 277-287.

TIEDE, D., LANG, S. and HOFFMANN, C., 2008 Domain-specific class modelling for one-level representation of single trees. In *Object-Based Image Analysis - Spatial concepts for knowledge-driven remote sensing applications*, T. Blaschke, S. Lang and G. Hay (Eds.), pp. 133-151 (New York: Springer).

VAUHKONEN, J., ENE, L., GUPTA, S., HEINZEL, J., HOLMGREN, J., PITKÄNEN, J., SOLBERG, S., WANG, Y., WEINACKER, H., HAUGLIN, K. M., LIEN, V., PACKALÉN, P., GOBAKKEN, T., KOCH, B., NÆSSET, E., TOKOLA, T. and MALTAMO, M., 2011 Comparative testing of single-tree detection algorithms under different types of forest. Forestry: An International Journal of Forest Research. **doi: 10.1093/forestry/cpr051**

WANG, Y., WEINACKER, H., KOCH, B. and STERENCZAK, K., 2008 Lidar point cloud based fully automatic 3d single tree modelling in forest and evaluations of the procedure. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37 (part B6b), Beijing, China, pp. 45–52.

WEINACKER, H., KOCH, B., HEYDER, U. and WEIANCKER, R., 2004 Development of filtering, segmentation and modelling modules for lidar and multispectral data as a fundament of an automatic forest inventory system. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (part 8/W2), Freiburg, Germany, pp. 50-55.

**About the presenting author:**
Since the completion of doctoral work on single tree LiDAR remote sensing in European forest, he is working as a Post-doctoral Researcher in an EU project 'Flexwood' in the Department of Remote Sensing and Landscape Information Systems, Albert-Ludwigs University, Freiburg, Germany. At present, he is engaged in LiDAR based stand level forest inventory for large area mapping. He has several years of experiences in the applied remote sensing in corporate sectors.

**About the co-author of the paper:**
At present, she is the Head and Professor of Department of Remote Sensing and Landscape Information Systems, Albert-Ludwigs University, Freiburg, Germany. She is engaged in teaching and research activities of the department. She has been recognized with many reputed awards for her scientific contribution in the field of remote sensing and forestry. She has been the leading scientist and project leader of many national and international projects sanctioned from EU and German ministry and other scientific organizations.