TREE SPECIES MAPPING BY COMBINING HYPERSPECTRAL WITH LIDAR DATA

P. Kempeneers, F. Van Coillie, W. Liao, F. Devriendt, K. Vandekerkhove

VITO-TAP, Boeretang 200, B-2400 Mol, Belgium

1. INTRODUCTION

This study deals with data fusion of hyperspectral and LiDAR sensors for forest applications. In particular, the added value of different data sources on tree species mapping has been analyzed. A total of seven species have been mapped for a forested area in Belgium: Beech, Ash, Larch, Poplar, Copper beech, Chestnut and Oak. Hyperspectral data is obtained from the APEX sensor in 286 spectral bands. LiDAR data has been acquired with a TopoSys sensor Harrier 56 at full waveform. Confirming previous research [1], it has been found that airborne LiDAR data, when combined with hyperspectral data, can improve classification results. The novelty of this study is in the quantification of the contribution of the individual data sources and their derived parameters.

LiDAR information was combined with the hyperspectral image in a data fusion approach. Different data fusion techniques were tested, including feature and decision fusion. Decision fucsion produced optimal results, reaching an overall accuracy of 96% (Kappa [3] of 0.95).

2. AVAILABLE DATA

The study area was the Wijnendale in Belgium ($51°3'50$" N, $3°2'31''$ E). Tree heights ranged from 7 m to 39 m, with a median height of 26 m (see Figure 1(a)). LiDAR data was obtained from a TopoSys sensor Harrier 56 at full waveform. The study area was acquired in four different flight lines. The resulting point density was 13.81 m^{-2} with a point spacing of 0.27 m (using all returns).

Hyperspectral data was acquired with the Airborne Prism Experiment (APEX). The APEX instrument covers a wide spectral range, including the short wave infrared ($372 - 2498$ nm). After removing some noisy bands in the blue part of the electromagnetic spectrum, a total of 286 spectral bands were retained. The spatial resolution was 1.5 m. Radiance values were atmospherically corrected to top of canopy reflectance, based on the radiative transfer model MODTRAN4 [4]. Geometric correction was based on direct georeferencing according to [5].

Reference data for the tree species were obtained in the field. A total of 1450 trees were labeled for the seven species. The sample size was unbalanced, which reflects the actual situation for our forest plot (see Table 1).

3. METHODS

The LiDAR data were processed in LAS format and converted to a raster grid with a cell size equal to the spatial resolution of the hyperspectral sensor (1.5 m). Two products were obtained: a vegetation height model (VHM) and percentile height values (PHV). The latter was represented as a multi-band image of 11 bands.

Data obtained from the the hyperspectral sensor (APEX) were also available as a multi-band image with 286 spectral bands. Before conducting a data fusion on the two data sources, the images were first analyzed individually. A supervised classification algorithm was trained based on the reference data acquired in the field (see Table 1). For the classifier, a support vector machine

Table 1. Tree species reference data base

class nr	species	sample size
1	Beech	401
2	Ash	67
3	Larch	116
4	Poplar	416
5	Copper beech	80
6	Chestnut	67
7	Oak	303
	Total	1450

was selected, using a radial kernel function. The regularization parameter (C: cost or penalty parameter of the error term) and the kernel parameter γ were optimized with a grid search algorithm for each data source individually. The optimal bands of the respective data sources were selected using a sequential floating forward search algorithm [2]. For the objective function, we used the Kappa coefficient.

Two common approaches were tested to fuse the LiDAR and hyperspectral data. The first involved feature fusion, which stacked the features (bands) of both data sources (PHV + APEX). The stacked feature set was then fed into a support vector machine. The second approach is referred to as decision fusion. Two separate classifiers were first trained for each data source individually (PHV and APEX). The (soft) output of the classifier representing the posterior probabilities of the classes were then saved as the respective images prob_{PHV} and $\text{prob}_{\text{APEX}}$. These image were then used as input for a newly trained SVM, resulting in the final output (classes). The classified map was filtered with a sieving filter to remove objects smaller than five pixels.

4. RESULTS

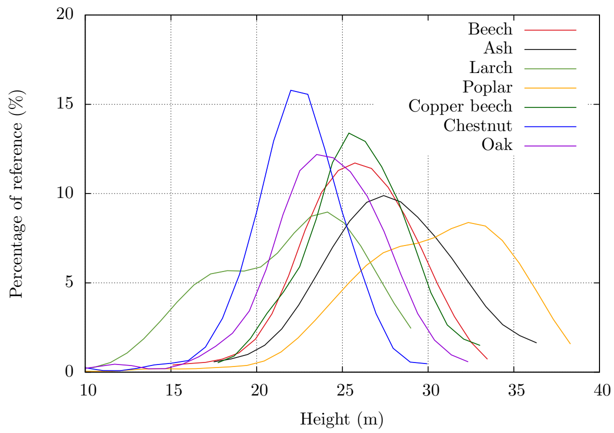
The results shown in Figure 1(b) represent the classification accuracies based on LiDAR data only. The accuracy is expressed by the Kappa coefficient [3]. As a reference, the horizontal line (Kappa=0.25) shows the accuracy using the single tree height feature (VHM). It is shown that the height profiles contain valuable information for the tree species classification. However, the maximum accuracy is relatively low (0.5) and is reached after four features. This indicates that the structure information derived from LiDAR contains valuable information, but is not sufficient for mapping the tree species under study. The height layers of the VHM that are selected by the feature selection algorithm are shown in brackets.

The results of the tree species classification are shown in Tables 2 and 3 and Figure 2(c). The results are based on a two-fold cross validation of the reference data. In Table 2 it is shown how the classification accuracy increases by including the different information sources. Column PHV represent the producer (PA) and user (UA) accuracies based on the percentile height values. The next two columns are based on hyperspectral dat only (APEX). The last two columns are based on decision fusion of the LiDAR and hyperspectral data sources. It is shown that decision fusion outperforms feature fusion. A confusion matrix is shown in Table 3 for the decision fusion only. Reference data are shown in rows and map data are shown in columns.

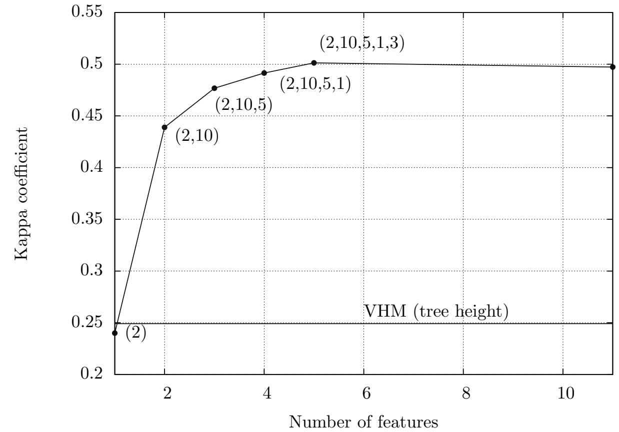
5. REFERENCES

[1] Michele Dalponte, Lorenzo Bruzzone, and Damiano Gianelle, “Tree species classification in the southern alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and lidar data,” *Remote sensing of environment*, vol. 123, pp. 258–270, 2012.

[2] Pavel Pudil, Jana Novovičová, and Josef Kittler, “Floating search methods in feature selection,” *Pattern recognition letters*, vol. 15, no. 11, pp. 1119–1125, 1994.

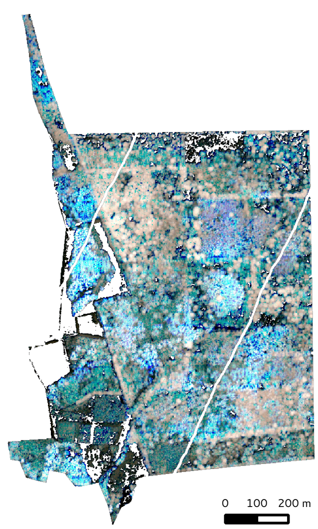


(a) Histogram of reference tree heights for different species



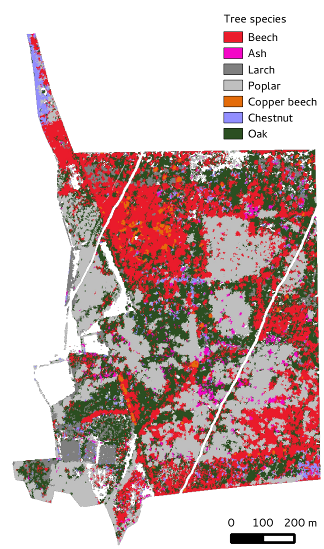
(b) Feature selection of percentage height values

Fig. 1. input data (LiDAR and hyperspectral) and tree species map for Wijnendale forest



(a) percentile height values at levels 2, 5 and 10

(b) APEX quicklook



(c) tree species map based on decision fusion

Fig. 2. input data (LiDAR and hyperspectral) and tree species map for Wijnendale forest

Table 2. Producer (PA) and user (UA) accuracies for different sources of data

class	species	PA	UA	PA	UA	PA	UA	PA	UA
		PHV		APEX		feature fusion		decision fusion	
1	Beech	83	55	79	78	86	78	96	96
2	Ash	0	0	55	88	48	80	91	98
3	Larch	23	71	80	85	83	82	99	99
4	Poplar	81	76	93	85	93	87	97	96
5	Copper beech	0	0	100	99	99	98	99	99
6	Chestnut	0	0	27	82	52	85	97	100
7	Oak	64	53	72	64	67	71	95	96

Table 3. Confusion matrix showin producer (PA) and user (UA) accuracies (Overall accuracy: 96.4, Kappa=0.95)

1	2	3	4	5	6	7	
1	384	1	0	8	1	0	7
2	2	61	0	4	0	0	0
3	1	0	115	0	0	0	0
4	5	0	0	405	0	0	6
5	1	0	0	0	79	0	0
6	1	0	0	1	0	65	0
7	8	0	1	5	0	0	289
PA	96	91	99	97	99	97	95
UA	96	98	99	96	99	100	96

[3] R. G. Congalton, "A review of assessing the accuracy of classification of remotely sensed data," *Remote sensing of environment*, vol. 37, pp. 35–46, 1991.

[4] A. Berk, GP Anderson, PK Acharya, JH Chetwynd, LS Bernstein, EP Shettle, MW Matthew, and SM Adler-Golden, "Modtran4 users manual," *Air Force Research Laboratory, Space Vehicles Directorate*, 1999.

[5] Q. A. Abdullah, *Manual of Photogrammetry*, chapter Photogrammetric platforms, pp. 677–730, American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland, USA, 5 edition, 2004.