

# ADABOOST-BASED FEATURE RELEVANCE ASSESSMENT IN FUSING LIDAR AND IMAGE DATA FOR CLASSIFICATION OF TREES AND VEHICLES IN URBAN SCENES

Yuzhang Wei <sup>a</sup>, Wei Yao <sup>b\*</sup>, Jianwei Wu <sup>c</sup>, Michael Schmitt <sup>a</sup>, Uwe Stilla <sup>a</sup>

<sup>a</sup> Photogrammetry and Remote Sensing, Technische Universitaet Muenchen, Arcisstr. 21, Munich, Germany

<sup>b</sup> Dept. of Geoinformatics, Munich University of Applied Sciences, Karlstr. 6, Munich, Germany

<sup>c</sup> College of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China

**KEY WORDS:** Imagery, LiDAR, trees, vehicles, classification, urban areas

**ABSTRACT:** In this paper, we present an integrated strategy to comprehensively evaluate the feature relevance of point cloud and image data for classification of trees and vehicles in urban scenes. First of all, point cloud and image data are co-registered by back-projection with available orientation parameters if necessary. After that, all data points are grid-fitted into the raster format in order to facilitate acquiring spatial context information per pixel/point. Then, various spatial-statistical and radiometric features can be extracted using a cylindrical volume neighborhood. Classification results as labeled pixels can be acquired from the classifier, and after appropriate refinements we obtain the objects of trees and vehicles. Compared to other methods which have assessed the classification and relevance simultaneously using a single classifier, we first introduce AdaBoost classifier combined with contribution ratio to provide both classification results and measures of feature relevance, and then utilize Random Forest classifier to evaluate and compare the feature relevance from a more independent viewpoint. In order to confirm the accuracy and reliability of classification and feature relevance results, we consider not only characteristics of the classifiers itself, but also errors of data co-registration and alterable parameters. We apply the procedure to two different datasets. In the dataset requiring co-registration a-priori, the AdaBoost classifier even achieves a great accuracy of 96.99% for trees and 83.45% for vehicles. The quantitative results of feature relevance assessment highlight the most important features for classification of tree covers and vehicles, such as NDVI, LiDAR intensity, planarity and entropy. By comparative analysis of the two independent approaches, the reliable and consistent feature selection for classification of trees and vehicles from LiDAR and image data could be validated and achieved, being unrelated to the classifiers.

## 1. INTRODUCTION

Urban scene classification is an important topic in the field of remote sensing. Recently, point cloud data generated by LiDAR sensors and multispectral aerial imagery have become two important data sources for urban scene analysis. Multispectral aerial imagery with high resolution provides detailed texture information, while point cloud data is more capable of presenting the geometrical characteristics of objects.

The current state of urban scene classification can be categorized by different objects, different data sources, and also different algorithms. During the last decade more papers referring to classification concentrated on specified objects. Many marvelous works (Jutzi & Stilla, 2003; Fauvel, 2007; Yao & Stilla 2010; Guo et al., 2011) have been done in extracting objects like buildings and roads, while trees and vehicles in urban areas are also interesting objects (Secord, 2005; Stilla et al., 2007; Höfle & Hollaus, 2010; Yao et al., 2011). However, classification of trees and vehicles might involve more complicated situations due to the various characteristics and appearances of the objects. More important, we want to know which features are relevant and how relevant they are with respect to the classification of specific objects. Moreover, we have to make sure the relevant features are reliable and consistent. Therefore, we intend to develop a comprehensive strategy of extracting relevant features from the two data sources for classification of trees and vehicles in different urban scenes while evaluating the reliability of the feature relevance. A framework for further researches in this field can be provided.

Straub (2004) gave a detailed description of an automatic approach on tree detection from image and LiDAR data, which shows the potential of 3D target modeling on trees in urban areas. Secord & Zakhor (2005) provided a two-step method for tree detection in co-registered aerial image and range data obtained via LiDAR consisting of segmentation followed by classification. Yang & Praun (2009) presented an automatic approach to classify trees from aerial imagery on pixel-level using one-class AdaBoost. Template matching is adopted followed by greedy selection to locate the trees. The strategy is quite parameter-sensitive using only image data, and without multiclass involved. Höfle & Hollaus (2010) combined the full-waveform LiDAR and image data to classify vegetation in urban scenes. Vehicles are also discussed but not towards integrating aerial imagery with LiDAR. Grabnera et al. (2008) used an online boosting-based strategy to detect cars solely from aerial images. Jutzi & Stilla (2003) presented a complete approach for classification of urban objects by laser pulse analysis. Syed et al. (2005) showed a complete automatic strategy of classification of different classes using both imagery and LiDAR data, which indicates the potential of complementation of the two data sources. Additionally, some papers have concentrated on the data source of LiDAR. For example, Fauvel (2007) used both spectral and spatial methods to classify objects from LiDAR data. Mallet et al. (2008) presented their works focusing on full-waveform LiDAR data to partition urban areas into building, vegetation, natural ground and artificial ground regions, however, the procedure also noticed by the authors could be improved in feature selection and classifier parameterization.

All the above works are excellent with great efforts, but none of them refer to feature relevance evaluation for classification in

---

\* Corresponding author: yao@hm.edu

urban scenes. Some researchers have developed approaches based on combination of both data sources for urban scene classification, but seldom assessed the feature relevance either towards vehicle detection or in the context of geometrical fusion of the two data sources. On the other hand, these two data sources sometimes are acquired in different mechanisms or at different time. Therefore, co-registration is often required in this case.

There exist already some well-developed algorithms for feature selection by theoretical analysis. Dash & Liu (2003) provided an approach based on consistency search for feature selection, which is original for the field of Artificial Intelligence. Guyon & Elisseeff (2003) introduced the variable and feature selection for machine learning. However, those algorithms were not frequently applied to quantifying the relevance in the task of classification, especially for the objects of trees and vehicles from image and LiDAR data. Still some researchers paid their attentions to this topic. Guo et al. (2010) focused on image and full-waveform LiDAR data, providing both classification results and measures of feature importance for classification of building, vegetation and ground surface. The research is comprehensive with detailed feature extraction and well-built classifier. However, only one single classifier is selected, the feature relevance results are subjective without validation. Also trees and vehicles are not specified as classes. As an update of their previous research, Mallet et al. (2011) indicated more details about the feature relevance in full-waveform LiDAR data. However, those relevance results have not proved to be unrelated to the classifiers and some of them are quite sensitive to the parameters and data.

In this work, we concentrate on the data sources of aerial multispectral imagery and LiDAR data, and using AdaBoost for classifying trees and vehicles and characterizing feature relevance by contribution ratio. However, those results are still assumed to be subjective without any verification. In order to assess the reliability and consistency of the feature relevance, we introduce the Random Forests classifier to generate classification results simultaneously providing feature relevance as its natural capability. The classification and feature relevance results from AdaBoost are expected to highlight the most relevant features for classification of trees and vehicles in urban scenes. By comparing with Random Forests, the feature relevance is verified towards the reliability and consistency across different classifiers. Therefore, in this paper, we develop a strategy to independently evaluate the feature relevance for classification of trees and vehicles in fusing airborne LiDAR and imagery. By comparing the results from different approaches, we can attain the possibility to prove the consistency and reliability of feature relevance.

Section 2 presents the theory and methodology of the classifiers and relevance assessment. Section 3 defines the used features. Section 4 describes the test datasets, data pre-processing, and how the experiments are carried out. Results are exhibited in section 5 with discussions made meanwhile. The conclusion will be deduced in Section 6.

## 2. CLASSIFIERS AND FEATURE RELEVANCE

### 2.1 AdaBoost

AdaBoost is the abbreviation of Adaptive Boosting, and it solves the problem of combining a bundle of weak classifiers to create a strong classifier which is arbitrarily well-correlated

with the true classification. The algorithm consists of iteratively learning weak classifiers with respect to a distribution and adding them to a final strong classifier. Once a weak learner is added, the data is reweighted according to the weak classifier's accuracy, examples that are misclassified gain weight and examples that are classified correctly lose weight. The Adaptive Boosting is formulated by Freund & Schapire (1999), who also introduce the original Boosting algorithm. AdaBoost is adapted to take full advantage of the weak learners. In this paper, we use an open-source AdaBoost toolbox with one tree weak learner CART, more details can be found in the reference.

AdaBoost contains two phases as well, namely training and prediction, in the training phase, it repeatedly trains a weak classifier of  $T$  rounds, where  $T$  is the number of the weak classifiers. As shown in the pseudo code:

```

Training Data as Input :  $(x_i, y_i), i \in [1, m]$ 
Initialize :  $W_i^1 = \frac{1}{m}, h_i^1 = 0;$ 
for  $t = 1 : T$ 
     $\epsilon_t = \sum_i W_i^t * I(h_i^t \neq y_i);$ 
     $h_t = \text{argmin } \epsilon_t;$ 
    if  $\epsilon_t \leq 0.5$ 
         $\alpha_t = \frac{\ln\left(\frac{1 - \epsilon_t}{\epsilon_t}\right)}{2};$ 
         $W_{i+1}^t = W_i^t * \epsilon^{-\alpha_t h_t y_i};$ 
    end
end

```

The  $(x_i, y_i)$  represent the training data and its reference, for example,  $x_i$  is a row of feature values of a point, and  $y_i$  is the class this point belongs to.  $m$  represents the amount of the training data. In each round, the AdaBoost algorithm selects out the proper threshold for each feature and updates the weight of each feature. The training data  $x_i$  that the classifier  $h_t$  identified correctly are weighted less and those that it identified incorrectly are weighted more. Therefore, when the algorithm is testing the classifiers on the  $W_{t+1}^i$ , it will select a classifier that better identifies those examples that the previous classifier missed.

The output of the training phase is a final strong classifier:

$$H(x) = \text{sgn}\left(\sum_{t=1}^T \alpha_t h_t\right) \quad (1)$$

Where the  $\text{sgn}$  function is defined as

$$\text{sgn}(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases} \quad (2)$$

Then in the prediction phase, it uses the strong classifier built in the training phase for classification. The weight  $\alpha_t$  can be an important factor of evaluating the feature relevance using contribution ratio defined by Masamitsu (2006), and we can quantify it by equation 3:

$$CR_p = \left[ \frac{\sum_{t=1}^T \alpha_t}{\sum_{i=1}^T \alpha_i} \cdot \delta_k(p) \right] \cdot 100\% \quad (3)$$

where  $p$  is one feature and  $\delta_k$  is the Kronecker delta.  $\delta_k(p)$  means that if the feature is chosen in round  $t$ , then Kronecker delta is 1 else equals 0. The range of  $CR_p$  is [0%, 100%]. In the experiments, AdaBoost is adapted to be capable of multiclass classification.

## 2.2 Random Forests

Random Forests is an ensemble classifier that is based on many decision trees and outputs the classification results as a combination of each individual tree. Prior to Random Forests, Boosting is supposed to be one of the best strategies for classification, but Random Forests can achieve a comparable accuracy while making an evident improvement in computational efficiency. The complete algorithm was firstly developed by Breiman (2001), and the basic idea came from random decision forests that was first proposed by Ho (1995).

The Random Forests method grows many classification trees, and each tree in the forest classifies the object based on an independently sampled random variable set. For classification, we put the variables downwards into each of the trees in the Forests. Then, each tree gives a 'vote' for that class. The algorithm chooses the classification which has the most votes. The whole procedure also contains two main parts: training and prediction. In the training phase, the algorithm generates  $T$  samples from the training data, and uses those samples to build Classification and Regression Trees (CART). Only a randomly selected subset of the input features is considered to split each node of CART. The variable that minimizes the Gini impurity is used for the split (Breiman, 2001). When the training set for a particular tree is drawn by sampling with replacement, about one-third of the cases are left out of the sample set, called out-of-bag (OOB) data. The OOB data is used to get a running unbiased estimate of the classification error as trees are added to the forest. It is also used to get estimates of variable importance from the training phase.

The importance of variable  $m$  can be estimated by randomly permuting all the values of variable  $m$  in the OOB samples for each tree, named mean decrease in accuracy, which is the difference in prediction accuracy before and after permuting variable  $m$ , averaged over all trees. Firstly, the difference between the numbers of votes for the correct class in the variable  $m$  permuted OOB data and the number of votes for the correct class in the untouched OOB data is calculated as equation 4:

$$Importance_{raw} = \frac{(Votes_{permuted} - Votes_{untouched})}{T} \quad (4)$$

Then, we divide the raw importance by its standard error to get a z-score and assign a significance level to the z-score assuming normality. This is the importance score for variable  $m$ . The range of the importance value is [0, 1].

The implementation codes of AdaBoost and Random Forests used in this paper are developed by Vezhnevets (2005) and Jaiental (2010), respectively.

## 3. FEATURE DEFINITIONS

In this paper, we combine point cloud and image data, and multispectral and intensity information are available but not always the case. In total 13 features are defined.

### 3.1 Basic Features

The so called basic features contain the features that can be directly retrieved from point cloud and image data.

- **R,G,B**: The three color channels of the digital image. As two data sets are used for experiments and one of them (named data set Vaihingen) provides color-infrared images, features  $R,G,B$  stand for infrared, red and green spectrums, But in the other

data set (Enschede), the features  $R, G,$  and  $B$  are normal bands of Red, Green and Blue. To avoid confusion, we always use the symbols  $R,G,B$  to indicate the three color channels of the image in order.

- **NDVI**: Normalized Difference Vegetation Index, defined as

$$NDVI = \frac{(NIR - VIS)}{(NIR + VIS)} \quad (5)$$

It can assess whether the target being observed contains green vegetation or not. This feature is specified for data set Vaihingen, because it provides color-infrared imagery.

- **Z**: The vertical coordinate of each point in the LiDAR data, as datasets used here are assumed to be flat.

- **I**: Intensity, which is provided by the LiDAR system for each point. The intensity is not available for data set of Enschede, since it provides the XYZ coordinates associated with color information.

### 3.2 Spatial Context Features

Based on the basic features, we intend to extract more features. Therefore, a 3D cuboid neighborhood is defined with help of a 2D square with radius of 1.25m in horizontal dimension as shown in Figure 1. All points located within the cell volume will be counted as the neighbors, the value 1.25m is chosen empirically.

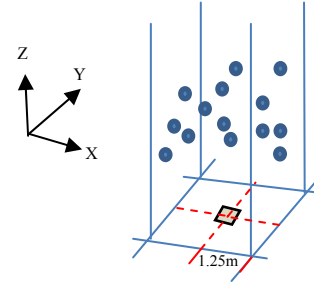


Figure 1: 3D cuboid neighborhood

-  $\Delta Z$ : Height difference between the highest and lowest points within the cuboid neighborhood.

-  $\sigma_z$ : height standard deviation of points within the cuboid neighborhood.

-  $\Delta I$ : Intensity difference between points having the highest and lowest intensities within the cuboid neighborhood.

-  $\sigma_I$ : Standard deviation of intensity of points within the cuboid neighborhood.

- **E**: Entropy, here being different from the normal entropy of images, we measure the entropy using intensities of the points within the cuboid neighborhood by equation 6 with  $K$  being the number of neighbors:

$$E = \sum_{k=1}^K \left[ (-I_k) \cdot \log_2^{I_k} \right] \quad (6)$$

The following two features  $O$  and  $P$  are based on the three eigenvalues of the covariance matrix from the XYZ coordinates of points within the cuboid neighborhood. The 3 eigenvalues  $\lambda_1, \lambda_2$  and  $\lambda_3$  are in descending order, and they can present the local tridimensional structure. This allows us to distinguish between a linear, a planar or a volumetric distribution of the points.

- **O**: Omnivariance, which indicates the distribution of points in the cuboid neighborhood. It's defined as

$$O = \sqrt[3]{\prod_{i=1}^3 \lambda_i} \quad (7)$$

-  $P$ : Planarity, defined as

$$P = (\lambda_2 - \lambda_3) / \lambda_1 \quad (8)$$

It has high value for roofs and ground, but low values for vegetation.

So far, in total we have defined 13 features if multispectral and LiDAR intensity information is provided besides RGB and XYZ information. However, as two different data sets are used for experiments, not all the features are always available.

## 4. EXPERIMENTS

### 4.1 Test Datasets

There are two different datasets. One of them was captured over Vaihingen in Germany with separate image and LiDAR data, while the other one from Enschede with image and LiDAR data integrated consistently.

**Dataset Vaihingen:** the digital aerial images of this dataset were captured using an Intergraph / ZI DMC on 24 July and 6 August 2008. The images are pan-sharpened color-infrared images with a GSD of 8 cm. Airborne Laserscanner (ALS) data was acquired on 12 August 2008 using a Leica ALS50 system with 45° field of view and a mean flying height above ground of 500m, the mean point density is 4 points/m<sup>2</sup>. Multiple echoes and intensities were recorded but due to the leave-on conditions, the number of points with multiple echoes is quite low, so no feature based on multiple echoes are generated. The test area is 145m×156m large in ground. As point cloud and image data were acquired at different times, the two data sources are post co-registered by geometrical back-projecting the point cloud into image domain with available orientation parameters. After that, all data points are grid-fitted into the raster format in order to facilitate acquiring spatial context information per pixel/point. We apply grid-fitting using an interval of 0.5m in ground ensuring that at least each resampled pixel can be allocated with one LiDAR point. Then, we obtain a resampled image of 326×291 pixels associated with 94866 LiDAR points. As color-infrared images and intensity information are provided, all 13 features are extracted for dataset Vaihingen:  $R, G, B, NDVI, Z, I, \Delta Z, \sigma_Z, \Delta I, \sigma_I, E, O$  and  $P$ . In the ground truth, 7590 pixels are generated as labels of trees while 1106 pixels are generated as labels of vehicles.

**Dataset Enschede:** which was acquired by the helicopter-based FLI-MAP 400 system from John Chance Land Surveys, Inc over Enschede in Netherlands in 2006. The FLI-MAP 400 system is a LiDAR sensor integrated with an additional line scan camera, which is able to provide true color values to each laser point (Red, Green and Blue attributes can be associated to the laser data points). Therefore, point cloud and image data are consistently integrated in this dataset. The test area covers the area of 153m×278m in ground but with irregular shape, and contains 1,197,686 LiDAR points integrated with RGB information. As only RGB and XYZ information are provided, in total 8 features are extracted for dataset Enschede:  $R, G, B, Z, \Delta Z, \sigma_Z, O$  and  $P$ . In the ground truth, 46893 points are generated as labels of trees while 10880 points are generated as labels of vehicles.

### 4.2 Design of Experiments

In this paper, AdaBoost and Random Forests are involved as two classifiers. We concentrate on the classification domain:

Others+Trees+Vehicles as 3 classes, which means that we will classify trees and vehicles, and for the other objects we consider them as Others. Two datasets are introduced for experiments, one from Vaihingen and the other from Enschede. In the experiments, we first use AdaBoost to generate classification results and then calculate the relevance results by contribution ratio. So far, we will get both classification and feature relevance results of AdaBoost, however, these results are one-sided without verification. Therefore, we apply Random Forests to generate both classification and relevance results simultaneously in order to do a comparatively analysis. For each dataset we can obtain two sets of classification results and two sets of feature relevance results. Results and discussions are presented in next section.

## 5. RESULTS AND DISCUSSIONS

### 5.1 Relevance Results

As explained in section 4.2, we use contribution ratio for AdaBoost to quantize the feature relevance and use Random Forests to generate comparative relevance results as Figure 2 and 3.

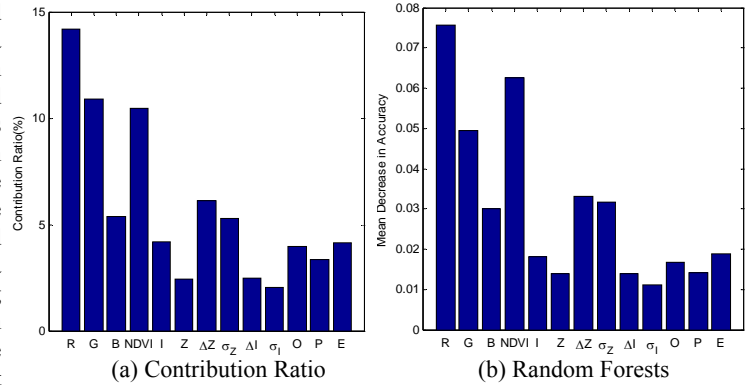


Figure 2: Feature relevance of dataset Vaihingen.

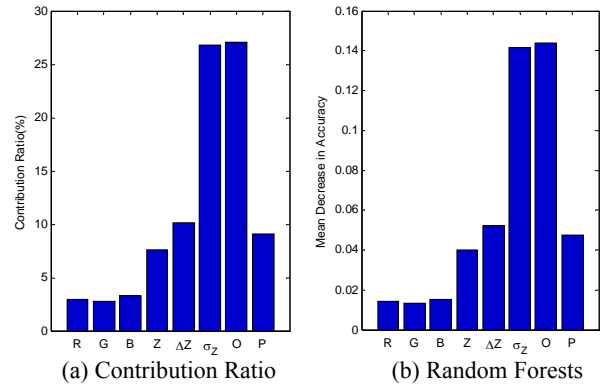


Figure 3: Feature relevance of dataset Enschede.

At first glance, we can say that the relative variations of feature relevance obtained by the two methods accord with each other in general. Moreover, we collect out the most relevant features (7 for Vaihingen, 4 for Enschede) indicated by contribution ratio of AdaBoost classifier for both datasets Vaihingen and Enschede in Table 1 in descending order from left to right.

Vaihingen	$R > G > NDVI > \Delta Z > \sigma_Z > B > E$
Enschede	$O > \sigma_Z > \Delta Z > P > Z$

Table 1: Most relevant features for datasets Vaihingen and Enschede

For dataset Vaihingen, the features  $R$ ,  $G$  and  $NDVI$  are the three most relevant features. The feature  $R$  stands for the first color channel of the color-infrared image, which indicates green objects in dataset Vaihingen.  $NDVI$ , which is derived by the combining spectral bands in the image of Vaihingen, is more relevant for classification of trees as expected. Compared to Guo et al. (2010), the features  $R, G, B$  in their research are sampled from normal visible spectral bands, and also  $NDVI$  is not available, so the features from images are not so important as here. On the other hand, the features  $\Delta Z$  and  $\sigma_Z$  also well contribute to the classification ranking as fourth and fifth places, respectively in Table 1, which agree with the results of Guo et al. (2010). Moreover, the ground truth shows that the amount of vehicle pixels is much smaller than trees. Therefore, all those factors could result in the high relevance of features  $R, G, B$  and  $NDVI$  for dataset Vaihingen.

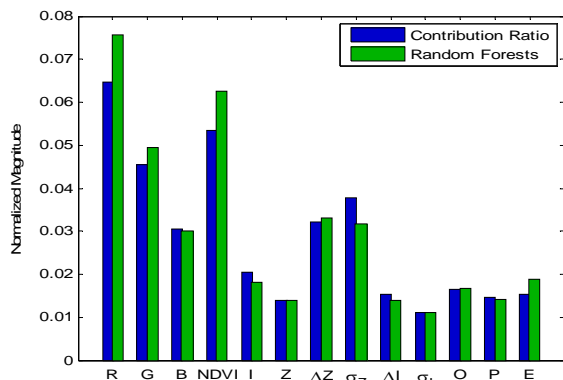


Figure 4: Feature relevance results of Contribution Ratio and Random Forests for dataset Vaihingen.

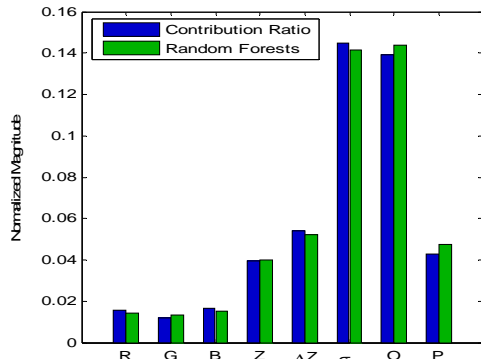


Figure 5: Feature relevance results of Contribution Ratio and Random Forests for dataset Enschede.

For data set Enschede, as shown in Table 1, features based on LiDAR data:  $O$ ,  $\sigma_Z$ ,  $\Delta Z$  and  $Z$  are the most relevant features, which can be easily assumed to be credited to the property of inherent alignment of image and point clouds in this dataset. However, features  $R, G, B$  result in much less importance, none of them take a place in the first four most relevant features. This is due to image sensitivity to the illumination angles and the presence of shadows in dense urban scenes. Vehicles are not frequently chosen as target objects in urban scene classification as trees. In works of Guo et al. (2010) and Mallet et al. (2011), buildings and ground surface are considered, but they are provided by exploiting more features based on full-waveform LiDAR technology. In order to further evaluate the consistency of relevance results, Figure 4 and 5 exhibit the relevance results from Contribution Ratio and Random Forests together with normalized magnitudes.

We can easily find out that the relevance results from the two different methods well agree with each other. The major differences happen in dataset Vaihingen in the first three most relevant features as  $R$ ,  $NDVI$  and  $\sigma_Z$ . For the less relevant features, the differences are almost negligible. Therefore, the feature relevance as a whole is assumed to be consistent and reliable for the two different methods and two different datasets.

## 5.2 Classification Results

In this section, we discuss the classification result of AdaBoost. Figure 6 and 7 show the classification maps of datasets Vaihingen and Enschede. As an overview, the results are very promising for trees in both datasets, whereas vehicles are worse classified in dataset Vaihingen. In Table 2 and 3, we summarize the classification accuracies for trees and vehicles.

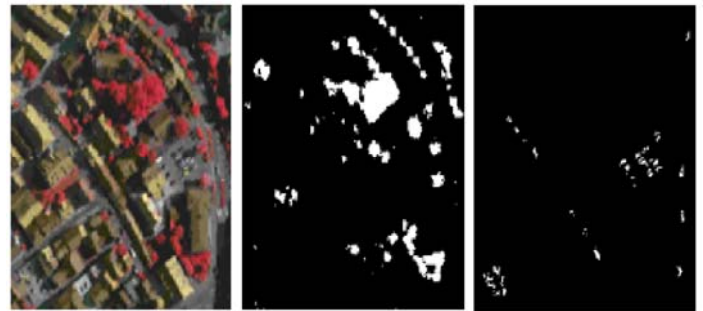


Figure 6: Classification maps of dataset Vaihingen with image(left), trees(middle) and vehicles(right).

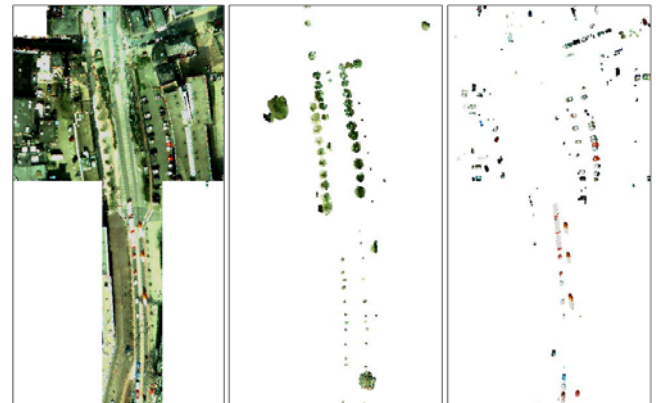


Figure 7: Classification maps of dataset Enschede with image (left), trees (middle) and vehicles (right).

	Commission Accuracy	Omission Accuracy
Trees	95%	91%
Vehicles	96%	60%

Table 2: classification performance for dataset Vaihingen

	Commission Accuracy	Omission Accuracy
Trees	97%	96%
Vehicles	94%	83%

Table 3: classification performance for dataset Enschede

From the accuracy results, we can see that trees are always well classified with the accuracies of  $>90\%$  for dataset Vaihingen and  $>95\%$  for dataset Enschede. For dataset Vaihingen, features  $NDVI$  and  $R, G, B$  contribute the most to the classification of trees while spatial context features seem not so important. This may be on

