in: Juliane Bendig und Georg Bareth (Hrsg.) (2014): Proceedings of the Workshop on UAV-basaed Remote Sensing Methods for Monitoring Vegetation. - Kölner Geographische Arbeiten, **94**. Köln: 21-32

TOWARDS UAV-BASED FOREST MONITORING

M. Drauschke[1], J. Bartelsen[1], P. Reidelstürz[2]
[1] Institute of Applied Computer Science, Bundeswehr University Munich, Germany
[2] Technologie Campus Freyung, Deggendorf Institute of Technology, Germany
Martin.Drauschke@unibw.de, Jan.Bartelsen@unibw.de, Patrick.Reidelstuerz@th-deg.de

Abstract

In this paper, we describe two experiments regarding the monitoring of a test site in the Bavarian Forest National Park using unmanned aerial vehicles (UAVs) and we show their results. In the first experiment, we show that it is possible to relatively orient the RGB images acquired by a small UAV in power glider configuration without any flight stabilisation and without integrated navigation system (INS) initial values. This enables a 3D scene reconstruction, i.e., we obtain a point cloud showing distinctive 3D points. A much denser point cloud showing trees with branches can be derived from dense image matching. In the second experiment, we demonstrate how multispectral imagery can be interpreted on demand, i.e., without producing an ortho-mosaic, but using reliable features and a powerful classifier. With our algorithm, we follow up the aim to detect bark beetle attack in an early infection stage in Sitka spruce, *Picea sitchensis*, in the Bavarian Forest National Park.

Keywords: unmanned aerial vehicle, image orientation, 3d reconstruction, random forest, detection of bark beetle attack in an early infection stage

1. Introduction

In 1713, Hannß Carl von Carlowitz firstly formulated the principle of *sustainability* applying it to the necessity of forest monitoring (von Carlowitz 1713). Today, 300 years later, the term "sustainability" is still significant and often used in politics to claiming actions regarding environmental research, e.g., forest monitoring, see also the project documentation (GAF AG 2009). Due to the easy access to remote sensing technologies or data, forest monitoring is a very active and interdisciplinary research field. While air- and even space-borne data can sufficiently be used for rough land cover classification, predicting biomass or recognizing vegetation damages at a forest scale, e.g., Fröhlich et al. (2013), Knuth et al. (2009) or White et al. (2007), close range remote sensing by airplanes flying at a low altitude and unmanned aerial vehicles (UAV) allows the analysis on tree-level, e.g., Carter et al. (1998) and Fassnacht et al. (2012). Analogue to precision farming, data showing details of individual trees can be analyzed for economic and ecological reason, e.g., with respect to plant growth or stress recognition.

So far, data acquisition using UAVs underlies the legal restriction of visual line of sight (VLOS). This means that the pilot must always see the UAV for confidently estimating its altitude, and it always must be possible to immediately switch into manual flight control mode in emergency cases. But in future, the operation restrictions of civil UAVs could be eased in regions with low hazard as forests, if the use of transponders in every aircraft would be regulated by law. Flight performances as flight time and flight distances of small UAVs in power glider configuration are not restrictive to date. Nevertheless, we test forest applications based on data acquisition by UAVs at exposed locations respecting the current VLOS rule. Thereby, we focus on two big challenges for an automatized UAV-based inspection of forests: (1) orienting the images for 3D reconstruction of the scene and for geo-referencing the data, and (2) classifying the scene by analysis of very complex and variable spectral signatures. Both points are addressed in this paper.

Normally, images from small and lightweight UAV flights are going beyond the scope of image orientation algorithms of classical aerial photogrammetry. Quite large off-image-plane rotation angles between the images and scale changes could be given in windy situations, even if established and proven autopilot systems are used, because of the low weight of such systems. Although we work with different autopilot systems, e.g., *Papparazzi* or *cloudcaptech piccolo*, we are also

interested in robust and rugged methods in relative image orientation, even if flights are carried out in manual radio controlled mode. I.e., images are taken from UAVs which are not stabilized by flight control systems, without camera triggering at planned positions and without integrated navigation system (INS) initial values as GPS/DGPS position and the three rotation angles from the autopilots inertial measurement system (IMU). Sometimes data acquisition could be very fast and very easy without the need of autopilot configuration and all the preparations for automatic or semi-autonomous flights. Consequently, if relative image orientation under these hindering conditions would be successful, it would be quite easy for autopilot assisted flights with initial values from INS. As first contribution of this paper, we show the robustness of our image orientation algorithm, its usability and potential, even for using aerial images from lightweight, not stabilized UAVs under windy weather conditions.

In this paper, we describe an experiment which is part of a project regarding bark beetle detection in spruces at an early infection stage. Further details on bark beetle damages in spruce forests and the research regarding their recognition see Niemann & Visintini (2005). We acquired multispectral imagery with a sports airplane in summer 2011. Three months later, we acquired RGB images as reference data by a UAV. In comparison to the airplane-based data acquisition, it is quite easy and cheap to capture UAV-based reference data. In the reference data we are able to visually recognize infected spruce trees in *red* or *grey attack stage* (the latter two of three infection stages). Three months before, these trees might be in *green attack stage* (first stage), which is not visually perceptible in optical light. So, as our second contribution, we show the potential of UAVs equipped with a multispectral camera for bark beetle detection in an early infection stage. Therefore, we employ a powerful multi-class classifier using reliable features which are partially based on previous work in precision farming applications.

In our flight campaign for obtaining reference data, we used a UAV in power glider configuration with 3.2 m wingspan and 3.6 kg maximum takeoff weight (MTOW) with long endurance performance in our campaign. As flight guidance system a *first person view* system (FPV) was used with video link in flight direction and onscreen display to show telemetry of flight data. The FPV system was equipped with GPS assisted auto tracking by two diversity patch antennas (Fig. 1, left). Hand launching and belly landing capabilities of the airframe allow its operation in semi rough terrain without a prepared runway. Using net landing the operation of such systems even would be possible in really rough terrains as wind throw areas (background in Fig. 1, right) without the disadvantages of the complexity of vertical takeoff and landing airplane capabilities. In the shown campaign 18 km flight distance were flown using less than 1/3 of the battery capacity. Using the FPV flight guidance system, orientation and navigation can be used to find relevant areas of bark beetle attacks.

This paper is structured as followed: In the next section, we describe how we orient the RGB images from the UAV flight obtaining a 3D reconstruction of the scene. This allows viewing the scene in stereo by constructing anaglyph images. They are used to manually inspect the scene and to derive ground truth for the scene classification described in section 3. We discuss our method

Fig. 1: Data acquisition at a test site in the Bavarian Forest National Park. Left: Preparing the UAV for its flight using a FPV flight guidance system. Right: UAV during its flight. Rough start and landing conditions as visible in the background can get avoided by using UAV platforms with hand launch and net landing capabilities as the shown power glider.

in section 4 and formulate tasks which should be considered in further research for forest monitoring using UAVs. In our last section we summarize our contribution.

2. 3D Reconstruction of the Scene

Reconstructing a scene from images means to orient the images relatively to each other, i.e., the camera poses are estimated, and to derive 3D points from subpixel positions of distinctive image points and the estimated camera poses. As described in detail by Hartley, R. & Zisserman, A. (2003) the change of the camera pose between two images can be efficiently estimated considering corresponding image points, which show identical object points in different images. For finding such corresponding points in image pairs, the matching must be robust to the following five problems: the images differ in (1) their viewpoint, i.e., an observed object is viewed under different perspectives (2) in their brightness and contrast, and (3) their scale, i.e., details of an observed object may not be recognizable in all images. Furthermore, (4) occlusions by other objects or (5) repetitive patterns may cause mismatches.

All five problems occur when a forest scene from images acquired by an UAV should be reconstructed in 3D. Especially UAVs in (motor) glider configuration often change their viewing direction from image to image within an acquired sequence. The ground resolutions of the acquired images differs a lot from the resolution of tree tops. Furthermore, the image scales also may vary, if the UAV was navigated at different altitudes. An additional sixth problem may also occur: the trees slightly move with the wind, i.e., corresponding image points may show the same object points, but the 3D coordinates of the object have moved between both images. We only consider small moves of the vegetation, because we only use images acquired on the same day. Large movements or deformations may occur, if a couple of weeks elapse between different image sets. The Bundler software Snavely et al. (2006) is often used for scene reconstruction from images. Although of its impressive results, Bundler has its limitations with respect to view point, scale and illumination changes, due to the fact that it uses SIFT matching Lowe (2004), which is focused on short baseline image sets. We employ the approach of Bartelsen et al. (2012) which is capable to orient unordered, highly resolved, wide baseline images.

The approach of Bartelsen et al. (2012) obtains distinctive image points employing scale space maxima (Lindeberg 1998). These key points consist of four parameters: the position in the image (x,y), the in-image-plane rotation, and the scale space maxima. The latter two define an image patch around (x,y), which is used to find corresponding image points by normalized cross correlation and affine least squares matching. According to Nistér (2004) five pairs of corresponding points are used to determine the relative camera poses in a RANSAC framework Fischler & Bolles (1981) to be robust even against a high percentage of mismatches. The solution is extended to image triplets (Torr & Zisserman 1997), which are combined to simultaneously determine the camera poses for all images. A robust bundle adjustment ensures that the approach converges with camera poses near the optimum. As result, Bartelsen et al. (2012) determine 3D points in space with a very low uncertainty. For further technical details also see Mayer et al. (2012).

We reconstructed the test site in the Bavarian Forest National Park from a UAV flight in power glider configuration. We considered 289 images acquired by a standard consumer RGB camera. Fig. 2 shows the result from relative image orientation by (Bartelsen et al. 2012), i.e., we show a sparse point cloud of about 530.000 3D points, each colored as in a corresponding image, the estimated camera positions in red and the links between the image triplets in green. Therefore, the trajectory of the UAV can be recognized well. If the UAV moved smoothly, the images could be linked in sequence. In contrast, sequentially acquired images are not linked, if the UAV flew a tight turn. Then it may occur that neighbored images in the sequence have opposite viewing directions, i.e., no correct corresponding points can be found in such image pairs because of the small or lacking overlap. Links between images from different sequence parts show that the approach of Bartelsen et al. (2012) performs well with images of different views and of different scale. Focusing on the reconstructed 3D points of the scene, Fig. 2 only provides little information. The paths are well recognizable due to their bright appearance. We are also able to distinguish spruce trees with green needles from the trunks of uprooted trees. This manual differentiation mostly relies on the color information of the 3D points, not by their position.

If the camera poses are known, the semiglobal matching by Hirschmüller (2008) could be applied to obtain the 2.5D representation of the scene and a dense 3D point cloud. Here, the algorithm realizes pixel wise matching. The search is limited to epipolar lines which are derived from the camera poses. The resulting dense point cloud already contains millions of 3D points, if the images have a resolution of five mega pixel. In comparison, during camera pose estimation only a few hundred 3D

points are reconstructed per image triplet. Consequently, many details are visible in the dense 3D point cloud, while the sparse one only shows little information of the scene.

Fig. 3 shows the dense 3D point cloud resulted from (Hirschmüller 2008). Because of its details we are able to recognize and assess the shape of individual trees. Both approaches were performed automatically. The reconstruction of the sparse point cloud was calculated in couple of hours, the dense matching was performed on a cluster in a few hours. Scale and quality of the used images based on lightweight UAV flights can be recognized in Fig. 4 (stereo model, anaglyph).

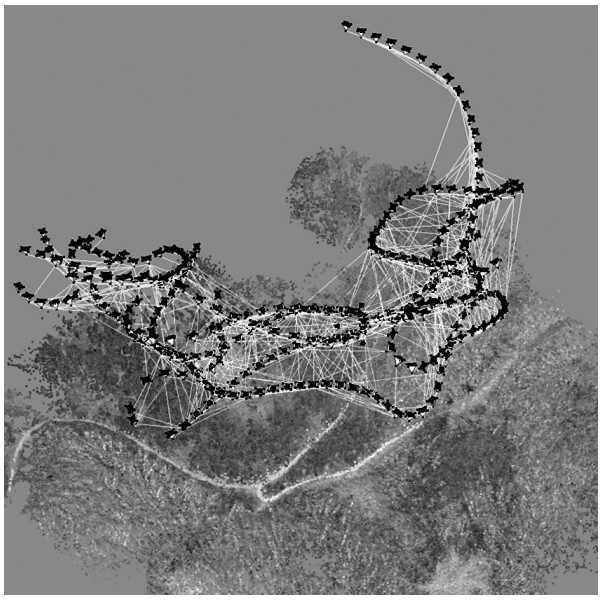

Fig. 2: Reconstructed forest scene with visualized camera positions (black) and linked image triplets (white).

3. Interpretation of Multispectral Imagery

In this section, we describe a procedure for automatic image interpretation of the forest scene. We acquire the multispectral images with the narrow band camera *TetraCam Mini MCA 6* which has a weight of 700 g. It is possible to carry such a camera in lightweight UAVs under 5 kg (MTOW). But for this work, we acquired multispectral imagery with a sports airplane on 5th August 2011. Onboard vibrations cause similar image blurring as if we would have flown the camera by a UAV. Furthermore, the trajectory of the airplane is similar to a UAV in (motor) glider configuration, i.e., we might have large off-image-plane rotations between consecutively acquired images. The sports airplane flew about 400 m over ground, so we could have obtained images with larger ground resolution, if we would have used a UAV.

The camera obtains six monochromatic intensity images with narrow-bands according to Tab. 1. Among the six filters we chose four filters of infra-red, which often have been successfully applied in precision farming, e.g. Zarco-Tejada et al. (2004). Consequently, we relinquished a filter for blue, and hence, the possibility to show normal colored images.

All six images are calibrated and show the same scene only with a very small translation. Having images taken

Fig. 3: Dense 3D point cloud showing standing and uprooted spruce trees. Blue colored regions show parts of the scene which were not visible from aerial perspective.

Fig. 4: Scale and image quality shown as stereo model (anaglyph).

band	b1	b2	b3	b4	b5	b6
wavelength [nm]	550 ± 5	670 ± 5	710 ± 5	780 ± 5	900 ± 10	950 ± 20

Tab. 1: Spectral filters for the acquisition of narrow-band intensity images.

with a focal length of 35 mm and sensor fields of 6.66 × 5.32 mm^2, we achieve a ground resolution with a pixel area of about 6 × 6 cm. The images were acquired on a cloudy day at Stubenbachtal, our test site in the Bavarian Forest National Park. Thus, the images show only small variations of reflection, i.e. no bright areas from sunlight and no shadows. About three weeks before our image acquisition, a tornado storm had raged at the particular forest site. Therefore, our test area does not only show young and mature Sitka spruce trees (*Picea sitchensis*) without any bark beetle (*Ips typographus*) damages, spruces of all three stages of attack (green, red and grey), and beech trees. The test site also contains many uprooted trees with water stress caused by the storm. We show a part of our scene in Fig. 5.

Our image interpretation is based on supervised classification, i.e., we train a powerful classifier by information from manually selected image parts. This data is obtained from a control flight by a UAV acquiring only RGB images, which are relatively oriented (Bartelsen et al. 2012). The 3-dimensional visualization by derived

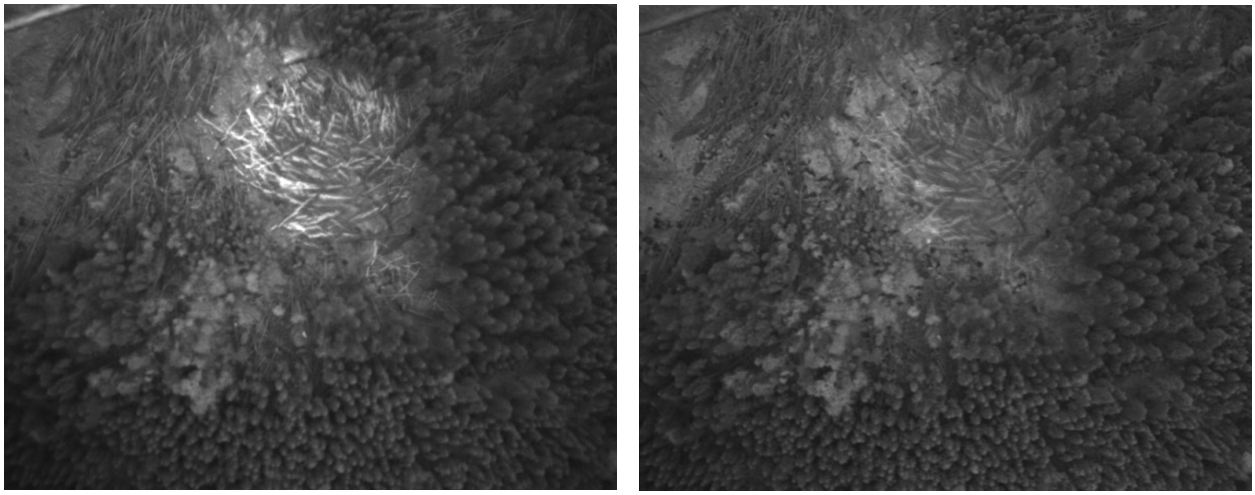

Fig. 5: Scene from Stubenbachtal, the monitored site in the Bavarian Forest National Park. Left: CIR image composition with bands (b4,b2,b1). Right: Color image composition of the remaining three bands (b3,b5,b6). In the center of the image, the scene shows lying peeled stems, which were caused by a storm a few weeks before data acquisition. In the upper left image quarter, there still lie many unpeeled stems. Below them, the scene shows beeches. Along the image bottom, many young spruces are visible, older and healthy trees stand at the right image border, and in the upper right quarter many trees in red attack stage can be found, and many trees of green attack stage stand between them.

anaglyph images (Fig. 4) enabled the manual interpretation of trees according to the aerial forest state inventory (Hildebrandt 1992). This forest state inventory is used to inspect single spruce trees, and to manually classify them according to the geometric structure of the tree's crown.

For annotation, we mark eleven classes in the scene. These classes are defined to represent homogeneous groups of objects. After applying the classifier, the results of some of these classes are merged together. We have defined and annotated the following classes:

1 *spruce-0-1*: This class comprises all healthy, well-formed spruces, i.e., trees with a crown formed by a pointed cone, showing a dense, regular and smooth coverage of needles into all directions (states 0 and 1 of inventory).
2 *spruce-2-3*: This class comprises all damaged spruces or trees having stress, i.e., trees with a crown formed by a cone with branches arranged like spokes, showing a sparse coverage of needles into many directions (states 2 and 3 of inventory).
3 *spruce-4*: This class comprises spruces, which are obviously dead (state 4 of inventory).
4 *spruce-young*: This class comprises all young spruces, i.e. trees, which do not show full crown yet. Later, this class is merged with *spruce-0-1* yielding one class for young or well-grown trees.
5 *beech*: This class comprises all leaf trees, primarily beeches.
6 *ground-plants*: This class comprises all other vegetation, mostly near the ground such as ferns, shrubs and bushes. Later, this class is merged with *beech* yielding one class for leaf vegetation.
7 *lying-trees* and *peeled-wood*: Both classes are later merged with *spruce-4* to the new class *wood*, because all three classes show wood, but no vital vegetation.
8 *soil, path, water*: All these classes form later the class *background*.

In the following subsections, we describe the classification technique, we explain which features are used for the classification, and we demonstrate the performance of our classification by experiments.

3.1 Random Forest

We implemented a random forest (RF) for classifying the multispectral imagery. Based on the formulation of the algorithm by Shotton et al. (2008) it is an efficient and effective classifier for multiple classes. It has been recently applied to multispectral image interpretation by, e.g., Tokarczyk et al. (2012) and Fröhlich et al. (2013). Instead of deriving a single decision tree, several uncorrelated, randomized decision trees are learnt and combined to achieve a much better performance.

For each single tree, a subset S_0 of the training data is randomly selected. Then several hierarchically ordered binary decisions are learnt to gradually split this subset into further subsets S_n^L and S_n^R, where n, $n \geq 0$, is the index of a tree's node. These decisions are coded in the inner nodes of the tree, and in the leaves, there are saved the class frequencies of the corresponding subsets.

The splitting of a subset stops, if either it consists of only one class, or its cardinality gets too small. Otherwise, S_n gets divided into the two subsets (Eq. 1)

$$S_n^L = \{s_i \in S_n | s_i(f_n) < \vartheta_n\} \text{ and}$$
$$S_n^R = \{s_i \in S_n | s_i(f_n) \geq \vartheta_n\} . \quad (1)$$

The parameter f_n is the selected feature and the parameter ϑ_n is the selected threshold, and both are chosen from randomly selected hypotheses for features and thresholds: the pair (f_n, ϑ_n) leads to the best split of subset S_n, i.e., the averaged entropy of the two split nodes is minimal with respect to the other hypotheses, see also (Shotton et al. 2008).

In the leaves of each randomized decision tree, the class frequencies of the corresponding subset are saved. These histograms are combined over all T trees, if new data gets classified. For a new sample, the corresponding leaf in each decision tree is determined. Then, we summarize the class frequencies over all T trees, the class with the maximum frequency is returned as classification result. The merging of some classes at the end of the classification is performed by replacing the particular class labels.

3.2 Reliable Image Features for Scene Interpretation

We derive our features directly from the acquired multispectral images. In opposition to other approaches, we do not determine an ortho-mosaic of the scene, where the images intensities are "balanced" to suppress illumination changes between the images (Miller 2004, 986). If ortho-mosaics are only calculated for geo-referencing the scene, this can also be done manually with the reconstructed scene. We train our classifier with data from several ten thousand pixels showing objects with dif-

ferent perspectives, resolution and illumination. So, we obtain a classifier which stably performs under such changes.

In land cover classification and precision agriculture, there have been studied several vegetation and chlorophyll indices, beginning with the normalized difference vegetation index (NDVI) by Rouse et al. (1974). Besides the three commonly known ratio values Green-to-NIR (b_1/b_4), Red-Edge-to-NIR (b_3/b_4) and the water index (b_5/b_6), we focused on indices which have been successfully applied in previous works, e.g., on water stress detection. The Green NDVI is similar to NDVI, but uses green light instead of red. Furthermore, we focused on the following vegetation and chlorophyll indices: SAVI, TVI, MTVI-1 and MTVI-2 as defined in Haboudane et al. (2004), TCARI, OSAVI and the ratio of both as defined in Haboudane et al. (2002), GCI and RECI as defined by Viña et al. (2011), and REIP as proposed by Guyot & Baret (1988). In the works above, sometimes hyperspectral data with slightly different wavelengths are used. Then we chose the spectrally closest filter of our camera, e.g., SAVI with 780 nm instead of 800 nm. REIP is originally defined using information of 740 nm, hence we interpolated the intensity values of 710 and 780 nm. If an index consists of a quotient of observed intensity values, it is stable regarding different illumination and reflectivity. Furthermore, indices like REIP and SAVI are stable with respect to soil and hydrological changes.

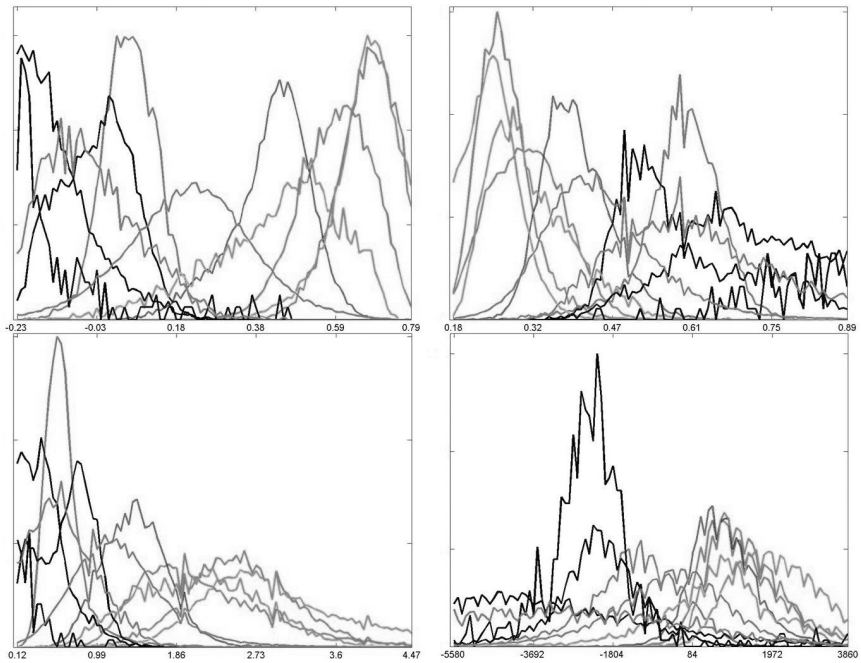

Fig. 6: Likelihood functions of the four indices (f.l.t.r) NDVI, G-to-NIR, GCI and TVI. The colors of the likelihood functions correspond to the following classes: green – spruces (state 0-1 or young), red – spruces (state 2-3), magenta – dead spruces or wood, cyan – beech, and black – the rest.

In Fig. 6, we show the likelihood functions of the eleven annotated classes regarding the four selected indices NDVI, G-to-NIR, GCI and TVI. The green curves show the likelihood functions of healthy spruces which are either young or trees in state 0 and 1 (inventory). The red curves show the likelihood functions of spruces in state 2 or 3 (inventory). In feature space of NDVI or the green-to-NIR ratio, both classes can get classified well. Using TVI, it seems that both classes cannot be discriminated. In feature space of GCI, all classes showing vital vegetation have similar likelihood functions. Thus, we need to combine several indices to construct a high-dimensional feature space for obtaining a reliable classification.

Since the vegetation and chlorophyll indices often show different likelihood functions for different classes, they build a good base for reliable features. Nevertheless, some indices, e.g., NDVI, SAVI and OSAVI are almost identical features, i.e., the equation 2

$$(1 + L) * \frac{b_4 - b_2}{b_4 + b_2 + L} \quad (2)$$

describes the NDVI, if $L = 0$, SAVI, if $L = 0.4$, and OSAVI, if $L = 0.16$. The parameter L in the equation above only causes a shift of the feature. Having a threshold-based classification scheme as we have implemented it for our random forest, these three indices have a similar performance, because a good threshold of one index can easily get transformed to become a good threshold of one of the other two indices.

Instead of tuning or inventing further vegetation indices like Fassnacht et al. (2012) to be invariant to various conditions, e.g., weather or soil, we applied strategies from machine learning. For instance, Guyon & Elisseeff (2003) recommend to arithmetically combine existing features, e.g., the observed intensity values of different channels or simple vegetation indices, to expand the feature space by generating additional features. Others, e.g., Tokarczyk et al. (2012) or Fröhlich et al. (2013) advise the importance of context information for obtaining a reliable classifier.

Consequently, we construct the following 104-dimensional feature space. The first six features are the intensity values of the observed spectral ranges. Then, we obtain 15 ratio-combinations of six bands, i.e., the ratio green to red, green to red edge, green to NIR etc. Furthermore, there are 15 combinations of the NDVI, since we have 15 different pairs of filters. Further 60 features are derived from combinations of three different bands, i.e., the three intensity values of the bands b_i, b_j and b_k (Eq. 3) form the features

$$\frac{b_i - b_j}{b_k + b_j}, \qquad \frac{b_j - b_i}{b_k + b_i}, \qquad \frac{b_i - b_k}{b_j + b_k} \quad . (3)$$

As last, we selected nine features some vegetation and chlorophyll indices which are not considered so far, and which might have a good impact on the classification: TCARI, OSAVI, and the ratio of both (defined according to the formulas by Haboudane et al. (2002)), TVI, MTVI-1 and MTVI-2 (defined according to the formulas by Haboudane et al. (2004)), and two implementations of REIP similar to the proposed formula in Guyot & Baret (1988), but with interpolated values for the intensity of 740 nm.

In Fig. 7, we visualize the values of each of the 104 dimensions of our feature space regarding the scene in Fig. 2. Since almost all features have a different range than typically intensity values have, we transformed each feature to 8-bit gray values. Therefore, we determined the 2- and 98-quantiles of a feature's values in the scene, and shifted those values to 0 and 255. Obviously, our feature space contains a lot of features with different characteristics, thus we might have reliable features for discriminating the classes.

3.3 Results

We have learnt $T = 200$ decision trees for the RF classifier. If T is much smaller, the performance of the classier decreases significantly. If a larger T is chosen, the performance does not really improve any more. We performed a cross validation with ten tests. In each experiment, we trained the classifier with 5000 randomly selected pixels of each of the 11 annotated classes, and we also chose 5000 other pixel per class for testing. After classification by RF, we merged spruces of state 0-1 and young spruces to one class, dead spruces and lying wood to one class, beech and ground plants to one class, and ground, water, paths to the background class. This class merging simplifies the visualization of the results, especially it shrinks the size of the confusion matrix in Tab. 2.

The confusion matrix has the following structure: Each row shows the classification results of samples belonging to this class, i.e., pixels belonging to class *spruce-2-3* are correctly classified with a probability of 0.755, and in 9.9% and 12.3%, these pixels are misclassified as *spruce-0-1* and *background*, respectively. So, the bold entries on the diagonal show the probabilities for correct classifications. Furthermore, if all entries of one column are summed up, the confidence of a classification can be determined. Hence, the probability is 0.817 for that a pixel classified as *spruce-2-3* really belongs to that class.

Totally, the quantitative analysis results in a recognition rate of 82.1%, i.e., more than four of five annotated pixels are correctly classified. The misclassification between the two classes *spruce-0-1* and *spruce-2-3* is

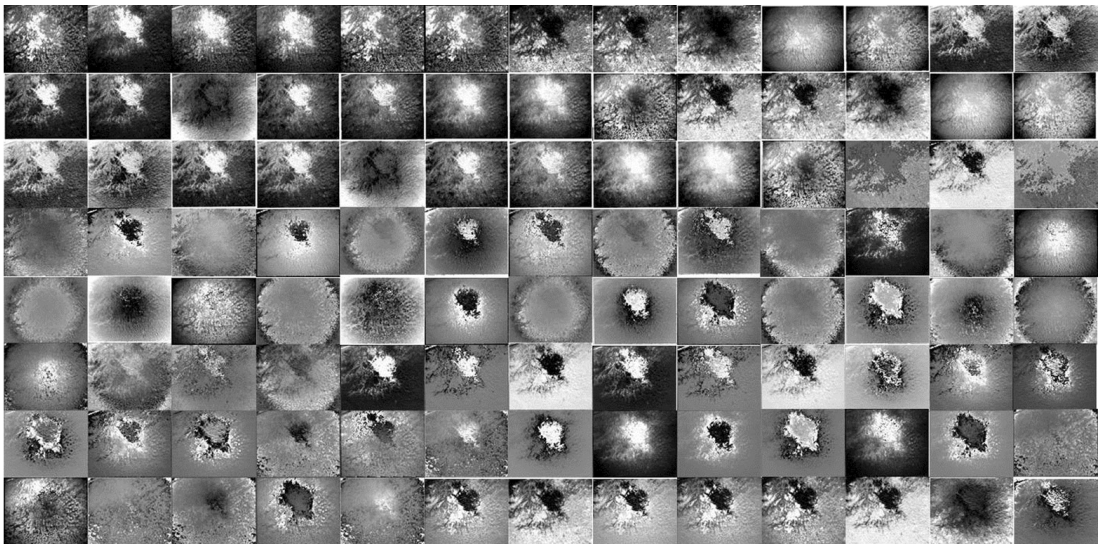

Fig. 7: Grey value visualization of all 104 features regarding the scene in Fig. 5. Each dimension of the scenes feature space has been linearly transformed to a gray value image.

	spruce-0-1	spruce-2-3	wood	beech	background
spruce-0-1	0.822	0.083	0.000	0.093	0.002
spruce-2-3	0.099	0.756	0.004	0.018	0.123
wood	0.000	0.004	0.823	0.000	0.173
beech	0.065	0.037	0.004	0.836	0.058
background	0.002	0.045	0.075	0.010	0.868

Tab. 2: Confusion matrix with five classes of classification with random forest with T = 200 trees.

relatively small. Thus, the discrimination between well-grown, possibly not infected trees and damaged, possibly infected trees works well.

Furthermore in Fig. 8, we visualize the classification results of all pixels of an image. The intensity images of this scene are given in Fig. 5. So, we can inspect interesting details of the scene. In the view's center, many lying trees are visible, and in this part of the image almost only wood and background have been recognized. In the upper left corner, many freshly fallen spruces are lying, here the classification shows many pixel being wood or background, but also damaged spruces were recognized. Here, the correlation between damaged trees and trees under water stress is visible. In the upper right corner, there stand many damaged spruces of state 2 or 3. And Fig. 8 also shows many pixels of this class at this site. In the bottom and at the right side of the image, many healthy and young spruces stand, and they were also well recognized. Here, the misclassification between the two classes *spruce-0-1* and *spruce-2-3* is also visible. Many pixels representing the top of a tree were wrongly classified as damaged tree.

4. Discussion

In our experiment regarding the reconstruction of the scene from UAV based images we demonstrated the current progress in this field. The relative orientation of images is not only possible if the flight has been conducted similar to an aerial photogrammetric data acquisition, i.e., the images in the acquired sequence do not have to be ordered, may have different scale and illu-

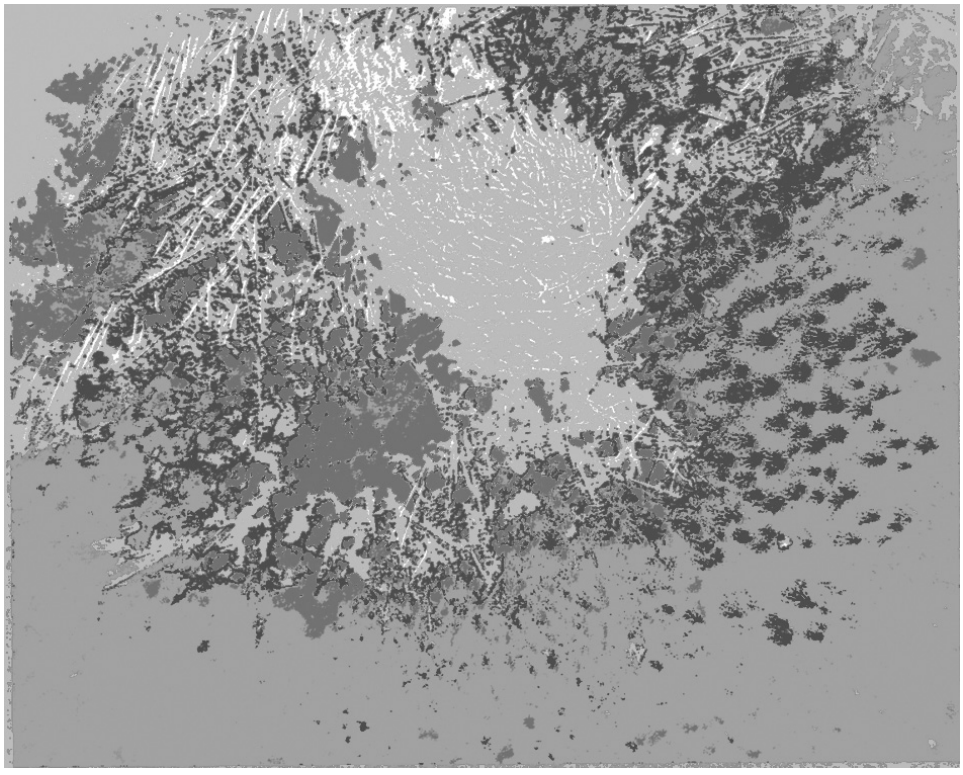

Fig. 8: Visualization of classification with random forest showing young spruces and mature spruces of state 0 or 1 of the aerial forest inventory in green, spruces of state 2 or 3 of the aerial forest inventory in red, dead trees and lying wood in white, beeches and ground plants in cyan, and background (soil, paths and water) in gray.

mination properties. Recent approaches like the one by Bartelsen et al. (2012) also work, if there are larger off-image-plane rotations between consecutively acquired images. Of course, even such robust approaches can fail, if the differences between images become too large.

Our second experiment regarding the scene classification leads to a more controversial discussion. To interpret the results, we point out that the defined classes *spruce-0-1* and *spruce-2-3* base on an manual inspection of 3-dimensional structures defined by the aerial forest inventory (Hildebrandt 1992), i.e., the trees were manually discriminated by the visual appearance of their crowns. Then we obtain from the classification that these trees can be discriminated using highly resolved multispectral information three months earlier. Of course, we cannot infer the cause of the damage, e.g., if it was water stress, or if the tree has been attacked by bark beetles. Maybe, this can be reached, if a Bayesian network is constructed where further information about properties of soil and climate are considered, or observed weather states are considered for the reasoning.

Besides the discussion on early detection of bark beetle detection, we want to point out two important issues of our approach. First, we do not rely on the interpretation of an ortho-mosaic, but use the original multispectral images. And secondly, we favor to generate many, often reliable features than tuning few good ones.

The direct use of the image for classification compared to traditional approaches with deriving an ortho-mosaic has a big advantage: We skip this part and, therefore, we do not need to balance the intensity value in the scene. Consequently, we have to select more training data to learn the classifier with many possible viewing directions and illumination conditions. The strong characteristic of the third dimension in forests might lead to large illumination variety. So we think of demonstrating our approach on simpler scenes, e.g., crop fields. Then we could show that the determination of ortho-mosaics is not necessary. This is part of our current and close future work.

Our feature space has a dimensionality of 104. For traditional classification methods, e.g., Maximum Likelihood classification assuming normally distributed features, this is already too big. Compared to recent works on land cover classification, e.g., Tokarczyk et al. (2012) or Fröhlich et al. (2013), we have a very small dimensional feature space. In such work, context information is used as feature which expands the feature space up to one million features. Such a big feature space is not a problem for RF classification, if enough good features are evaluated in the learning step at each splitting node. In each step, one-dimensional subspaces are considered. We also think of expanding our feature space by texture and contextual features, e.g., we should also include differences to neighbored pixels as features into our classification. Furthermore, we could postprocess the interpretation results by a majority vote in a pixel's neighborhood. This also leads to a smoothing of the classification output, i.e., isolated pixels will be removed in the classification output.

We are aware that we had relatively simple illumination conditions during image acquisition. Since it was a cloudy day, we do not have any strongly reflecting or shady image parts. Thus, our RF classifier has not learnt how the classes appear under more complex illumination conditions. Furthermore, we doubt that RF derived from observations at the Bavarian forest can successively be applied to other sites, e.g., the Black forest in south western Germany or forests in Canada. There, different soil properties and other spruce species could be found, and then a new RF classifier should be learnt from data of those places. We are confident that our RF is applicable at other time stamps as well. Each observed tree shows several age groups of needs, and therefore, the random forest has learnt these different signatures. Maybe, the random forest should get updated after a couple of years to include information of the last years.

We demonstrated that our approach is very successful regarding the classification of spruces labeled by 3D-inspection of their geometry using the aerial forest inventory (Hildebrandt 1992). So far, this classification has only been statistically analyzed as a snapshot, i.e., with almost no variation in time, place, weather and botanical diversity. Hence, we need to verify our results in additional studies. Primarily, we should analyze, if our close range remote sensing classification framework is transferable to classification of newly acquired images. Thereby, we should focus on seasonal changes, geological differences and on the employment of our approach at other sites.

Furthermore, we must evaluate the coherences between the 3D geometry of spruces and their bark beetle attack stages. In 2012, we already performed a small study with a continuous monitoring of our test site in the national park Bavarian forest. There, we weekly inspected individual spruces checking the state and degree of damage by bark beetles. Results of these experiments still need to be published.

5. Summary

We tested the capabilities of UAV for frequent short term forest monitoring. We showed that images from a UAV flight can be oriented, although there might be large off-image-plane rotation changes between consecutively acquired images. Therefore, all images must significantly overlap with other images of the flight. Scale differences and illumination changes are no problem, if they are not too big. We used the image orientation for deriving a stereo model which was used to manually inspect the scene, i.e., we manually interpreted the condition of single tree crowns using stereo view obtained from UAV based large scale images. This scene analysis was used to label multispectral images which we acquired with a sports airplane about three months before. This data was used to train a supervised classifier, random forest. In our tests on pixels which were not used for training we obtained a mean precision of 0.821. This shows that there are correlations between the phenotype (outer appearance) of spruce trees and its spectral signature. It also shows great capabilities of the classifier to detect different forest conditions using narrowband multispectral images. Finally, our whole approach advises great feasibilities and prospects of UAV based forest monitoring challenges.

Acknowledgements

The authors thank Dr. Heinrich Rall, former head of the department for research and documentation of the Bavarian Forest National Park, for his unlimited support of our work. Furthermore, we would like to thank the current national park authorities and the institute of robotics and mechatronics of the German Aerospace Center in Oberpfaffenhofen for their support.

References

Bartelsen, J., Mayer, H., Hirschmüller, H., Kuhn, A. & Michelini, M. (2012): Orientation and Dense Reconstruction from Unordered Wide Baseline Image Sets. Photogrammetrie – Fernerkundung – Geoinformation, **2012** (4): 421-432.

von Carlowitz, H.C. (1713): Sylvicultura Oeconomica Oder Haußwirthliche Nachricht und Naturmäßige Anweisung Zur Wilden Baum-Zucht.- Leipzig.

Carter, G.A., Seal, M.R. & Haley, T. (1998): Airborne Detection of Southern Pine Beetle Damage Using Key Spectral Bands. Canadian Journal of Forest Research **28** (7): 1040-1045.

Fassnacht, F.E., Latifi, H. & Koch, B. (2012): An Angular Vegetation Index for Imaging Spectroscopy Data – Preliminary Results on Forest Damage Detection in the Bavarian National Park, Germany. International Journal of Applied Earth Observation and Geoinformation **19**: 308-321.

Fischler, M. & Bolles, R. (1981): Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM, **24** (6): 381-395.

Fröhlich, B., Bach, E., Walde, I., Hese, S., Schmullius, C. & Denzler, J. (2013): Land Cover Classification of Satellite Images Using Contextual Information.- In: Sohn, G., Rottensteiner, F. & Wegner, J.D. (eds.): ISPRS Workshop on 3D Virtual City Modeling, 28.05.2013, Regina, Canada. ISPRS Annals II-3/W1:1-6.

GAF AG (2009): GMES Service Element Forest Monitoring (GSE FM) Final Report. Munich, Germany, 55 pp.

Guyon, I. & Elisseeff, A. (2003): An Introduction to Variable and Feature Selection. Journal of Machine Learning Research, **3** (Mar): 1157-1182.

Guyot, G. & Baret, F. (1988): Utilisation de la Haute Resolution Spectrale pour Suivre L'etat des Couverts Vegetaux.- In: European Space Agency (ed.): Proceedings of the 4th International Colloquium on Spectral Signatures of Objects in Remote Sensing, 17.-22.01.1988, Aussois, France. ESA Publications Division, 279-286, Noordwijk.

Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J. & Dextraze, L. (2002): Integrated Narrow-Band Vegetation Indices for Prediction of Crop Chlorophyll Content for Application to Precision Agriculture. Remote Sensing of Environment **81** (2-3): 416-426.

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J. & Strachnan, I.B. (2004): Hyperspectral Vegetation Indices and Novel Algorithms for Predicting Green LAI of Crop Canopies: Modeling and Validation in the Context of Precision Agriculture. Remote Sensing of Environment **90** (3): 337-352.

Hartley, R. & Zisserman, A. (2003): Multiple View Geometry in Computer Vision.- Cambridge.

Hildebrandt, G. (1992): Anwendungen der Fernerkundung zur Beurteilung des Gesundheitszustandes der Wälder. Walphot, Belgien.

Hirschmüller, H. (2008): Stereo Processing by Semiglobal Matching and Mutual Information. IEEE Transactions on Pattern Analysis and Machine Intelligence, **30** (2): 1582-1599.

Knuth, R., Thiel, C., Thiel, C., Eckhardt, R., Richter, N., Schmullius, C. (2009): Multisensor SAR Analysis for Forest Monitoring in Boreal and Tropical Forest Environments.- In: IEEE (The Institute of Electrical and Electronics Engineers) (ed.): Interna-

tional Geoscience and Remote Sensing Symposium, 12.-17.07.2009, University of Cape Town, South Africa. **V**: 126-129.

Lindeberg, T. (1998): Feature Detection with Automatic Scale Selection. International Journal of Computer Vision, **30** (2): 79-116.

Lowe, D. (2004): Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, **60** (2): 91-110.

Mayer, H., Bartelsen, J., Hirschmüller, H. & Kuhn, A. (2012): Dense 3D Reconstruction from Wide Baseline Image Sets.- In: Dellaert, F., Frahm, J.-M., Pollefeys, M., Leal-Taixe, L. & Rosenhahn, B. (eds.) Outdoor and Large-Scale Real-World Scene Analysis: 15th International Workshop on Theoretical Foundations of Computer Vision, 26.06.-01.07.2011, Dagstuhl, Germany. Lecture Notes in Computer Science **7474**: 285-304.

Miller, S. (2004): Photogrammetric Products.- In: American Society for Photogrammetry and Remote Sensing (pub.): Manual of Photogrammetry, 5th edition, 983-1013.

Niemann, K.O. & Visintini, F. (2005): Assessment of Potential for Remote Sensing Detection of Bark Beetle-Infested Areas During Green Attack: A Literature Review. Natural Resources Canada, 14 pp.

Nistér, D. (2004): An Efficient Solution to the Five-Point Relative Pose Problem. IEEE Transactions on Pattern Analysis and Machine Intelligence, **26** (6): 756-770.

Rouse, J.W., Haas, R.H., Schell, J.A. & Deering, D.W. (1974): Monitoring Vegetation Systems in the Great Plains with ERTS.- In: Freden, S.C., Mercanti, E.P. & Becker, M.A. (eds.): Third Earth Recources Technology Satellite Symposium, 10.-14.12.1973, Washington D.C., USA, 309-317.

Shotton, J., Johnson, M. & Cipolla, R. (2008): Semantic Texton Forests for Image Categorization and Segmentation.- In: IEEE Conference on Computer Vision and Pattern Recognition, 23.-28.06.2008, Anchorage, Alaska, USA.

Snavely, N., Seitz, S.M. & Szeliski, R. (2006): Photo Tourism: Exploring Photo Collections in 3D. ACM Transactions on Graphics, **25** (3): 835-846.

Tokarczyk, P., Montoya, J. & Schindler, K. (2012): An Evaluation of Feature Learning Methods for High Resolution Image Classification.- In: Shortis, M., Paparoditis, N. & Mallet, C. (eds.): XXII ISPRS Congress, Technical Comission III, 25.08.-01.09.2012, Melbourne, Australia. ISPRS Annals **I-3**: 389-394.

Torr, P.H.S. & Zisserman, A. (1997): Robust Parameterization and Computation of the Trifocal Tensor. Image and Vision Computing, **15** (8): 591-605.

Viña, A., Gitelson, A.A., Nguy-Robertson, A.L. & Peng, Y. (2011): Comparison of Different Vegetation Indices for the Remote Assessment of Green Leaf Area Index of Crops. Remote Sensing of Environment **115** (12): 3468-3478.

White, J.C., Coops, N.C., Hilker, T., Wulder, M.A. & Carroll, A.L. (2007): Detecting Mountain Pine Beetle Red Attack Damage with EO-1 Hyperion Moisture Indices. International Journal of Remote Sensing, **28** (10): 2111-2121.

Zarco-Tejada, P.J., Berjón, A. & Miller, J.R. (2004): Stress Detection in Crops with Hyperspectral Remote Sensing and Physical Simulation Models.- In: BruHyp Airborne Imaging Spectroscopy Workshop, 08.10.2004, Bruges, Belgium.

Contact information
Patrick Reidelstürz
Deggendorf Institute of Technology
Technologiecampus Freyung
Grafenauer Str. 22
D-94078 Freyung
Germany
Patrick.Reiselstuerz@th-deg.de
+49 (0) 8551 917 64 10