INVESTIGATING THE USABILITY OF UAV OBTAINED MULTISPECTRAL IMAGERY IN TREE SPECIES SEGMENTATION

Melinda Pap[1], Sandor Kiraly[2], Sandor Moljak[3]

[1] IoT Research Center, Eszterházy Károly University, Eger, Hungary - pap.melinda@uni-eszterhazy.hu
[2] Information Technology Department, Eszterházy Károly University, Eger, Hungary - kiraly.sandor@uni-eszterhazy.hu
[3] Innoregio Knowledge Center, Eszterházy Károly University, Eger, Hungary - moljak.sandor @uni-eszterhazy.hu

Technical Commission II

KEY WORDS: UAV, image segmentation, image processing, multispectral imagery, energy plants

ABSTRACT:

A widely used form of renewable energy are bioenergy crops. One form of it is the energy forestry that includes short rotation coppice plantations in which fast growing species of tree or woody shrub are grown (e.g. poplar, willow). The accurate prediction of forest biomass and volume can be used for the evaluation of plant breeding efficiency as well. The automatic tracking of plant development by traditional methods is quite difficult and labor intensive. Since energy forestries often contain different trees for estimating their volume it is essential to find segments containing the same tree species in the image.
We investigated the applicability of a low cost UAV and an intermediate cost UAV in the field of agricultural image segmentation that is the first stage of biomass estimation (Gatziolis et al., 2015, Gaulton et al., 2015).
This paper is a case study that shows the results of several segmentation algorithms applied on imagery obtained by a low cost UAV with low-cost camera, and imagery gathered by a UAV and camera set that are of higher quality and price. In the case study, we have observed two small forestry areas that contained six different tree species and their hybrids. Our results show that more expensive, better-equipped drone shots do not necessary provide significantly better segmentation.

1. INTRODUCTION

Plants can be divided into food, feed and industrial plants for cultivation purposes. Energy plants are important for preserving the Earth's ecology and as alternative energy sources like bio-fuel. They play an important role both in producing bio-fuel and heating electricity-generating power stations. Plenty of tree species (woody plants) can be planted as energy plants, for example Red oak (Quercus rubra), Gray Poplar (Populus canescens) or White Poplar (Populus alba).

To accurately estimate the volume and growth of the forestry, a detailed and precise three-dimensional (3D) representation of the area is crucial. Until recently, measuring the volume, spatial arrangement and shape of trees with precision has been constrained by technological and logistical limitations and cost. Traditional methods of plant biometrics provide merely partial measurements and they are rather labor intensive. Advances in Unmanned Aerial Vehicle (UAV) technology has made it feasible to acquire high-resolution imagery using lightweight cameras. Based on the imagery, three dimensional (3D) data can be constructed, that can be used for energy plants monitoring and assessing tree attributes (Mohan et al., 2017, Watts, 2012). The creation of a 3D model from 2D images obtained by UAV is necessary for measuring the volume of plants and monitoring their vegetation.

Structure from Motion (SfM) and multi-view stereopsis (MVS) techniques using digital cameras on small, low-cost UAVs are a potential cost-effective alternative for areas of woodland. SfM has emerged recently as a method for extracting the 3D structure of a scene from multiple overlapping photographs using bundle adjustment procedures (Fukunaga, Hostetler, 1975, Snavely et al., 2008, Westoby, Reynolds, 2012). They do not require the information on the 3D position of the camera or the 3D location of multiple control points, as opposed to traditional digital photogrammetric methods.

SfM can be used to generate high quality 3D point clouds for characterizing forest structures and can also be used to generate accurate Digital Surface Models (DSMs) from the 3D point clouds. It is the 3D representation of the surface of a terrain and DEM (Digital Elevation Model) which is a subset, and the most fundamental component of DSM (St-Onge, Véga, 2004, Cunliffe, 2016, Messinger, 2016). The DSM is essential in creating an orthophoto of the whole, scanned area. An orthophoto is a geometrically corrected uniform-scale photograph, so it is possible to use it for measurements.

Image segmentation is the process of separating or grouping an image into different image objects. An image object is a group of connected pixels in a picture, where the objects are homogeneous with respect to specific features. These features can be represented by RGB values, gray levels or textures, each exploring similarities between the pixels of a region. Other segmentation methods focus on finding boundaries between regions. There are many different ways of performing image segmentation, ranging from the simple thresholding method to different color image segmentation algorithms.

In this paper, we aimed to compare the usability of orthophotos acquired by two different UAVs, with different camera sets, for the purpose of image segmentation. We conducted this study in an energy forestry, with two study areas where hybrids of the same species are present, often with fuzzy borders.

2. DATA ACQUISITION

2.1 Study areas

Two sites were chosen for this study at the plantations of the Fleischman Rudolf Research Institute of the Eszterházy Károly University, at Kompolt, HU (47.738571 N, 20.234455 W, study area A. and 47.735889 N, 20.224807 W, study area B.), see Fig. 1. The first area is comprised of a variety of seven tree species which included Japanese willow, "Fertőszentmiklósi" willow, Raspolje poplar, Green maple, Ash, Poplar and five Bushy willow hybrids. In study area B, willows, acacias and six hybrids of poplar were present. Compared to the second area, the first one had a dense canopy structure making it suitable for evaluating the performance of SfM photogrammetry in closed canopies.

(a) Study area A (b) Study area B

Figure 1. The two study areas from Google Earth

The first aerial survey was conducted on 13 August 2017 using a DJI Phantom 2 quadcopter for both areas. Some results of the first survey have already been published. The second survey was conducted on 6 September 2018 using a DJI Inspire 2 quadcopter. The attributes and the parameters of the UAVs and the flights are presented in Table 1.

Parameter	Survey 1	Survey 2
UAV type	DJI Phantom 2 quadcopter	DJI Inspire 2 quadcopter
RGB Camera	GoPro Hero4, 12.1 Megapixel CMOS sensor, 4000 3000 resolution	Zenmus X5S, 20.9 megapixels, 4/3 inch CMOS sensor
Multispectral camera	no	4 multispectral sensors, capable of a resolution of 1.2 MP, and a 16 MP RGB sensor
Sun sensor	no	yes
IMU and magnetometer	no	yes
Integrated GPS	no	yes
Altitude	60 m	100 m and 70 m in the case of multispectral camera
Number of RGB images	185 and 245	139 and 263

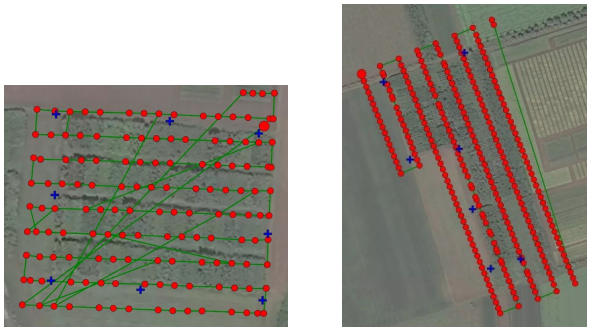
Table 1. Summary of the parameters of the two surveys: UAV and flight features

2.2 RGB images

During the first survey, the flight altitude was 60 meters at a speed of 5 m/s. The flight was planned with the DJI Ground Station software. The RGB camera automatically triggered (1 image / 2s) during the flight, capturing approximately an area of 14.42 x 10.77 m^2 with a pixel resolution of 3.5 mm in the case of both studied areas. We wanted to investigate the applicability of a low cost UAV and camera in the field of precisional agriculture. The drawbacks of such a drone and camera set are that the flight cannot be planned as accurately as for higher cost UAVs, meaning that instead of providing the flight planning software with the percentage of the wished overlap between consecutive photos, we can only set a timer for the exposure. Moreover, in this UAV setup, the GPS coordinates of the pictures were not accessible.

During the second survey, the flight altitude was 70 meters above the study area A, which was calculated from the launch point by the flight planning software. The drone covered the hand-selected area of approximately 180x164 meters at the set height and overlap in 9 flight lines. The overlap in the flight range was 90%, while the overlap between two neighboring lines was 80%. The total flight time was 6 minutes 38 seconds. During the flight 139 images were taken in orthogonal camera positions. The georeference was specified with 8 ground control points (see Fig. 2).

During the second survey, the flight altitude was 100 meters above the study area B, which was calculated from the launch point by the flight planning software. The drone covered the hand-selected area of approximately 326x433 meters at the set height and overlap in 7 flight lines. The overlap in the flight range was 90%, while the overlap between two neighboring lines was 75%. The total flight time was 13 minutes 21 seconds. During the flight, 263 images were taken in orthogonal camera positions. The georeference was specified with 7 ground control points (see Fig. 2).



(a) Flight plan of study area A (b) Flight plan of study area B

Figure 2. The flight path over study area A and B and Ground Control Points (GCPs)

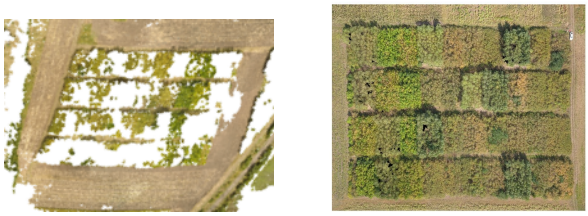
2.3 Multispectral images

During the second survey, the multispectral investigation of the study area B was performed at 70 meters and with a 70% overlap. The drone flew 8 flight lines at 6 m/s. The georeference was refined with 7 ground control points. During the flight 788 images were taken from 197 positions and in 4 channels (Green, Red, Red edge, NIR) in an orthogonal camera position.

2.4 Orthophoto and DEM

There are several 3D reconstruction software available but they perform the best on scenes that do not contain complex, non-convex structures. Our study area contains dense canopy as well as regions with sparse forest, where the camera could see below the trees. Such a surface is rather complex to construct by software. The aim of the 3D model creation was to obtain the orthophoto of the area, as well as the Digital Elevation Model (DEM). We have observed the applicability of two softwares in the process of 3D reconstruction.

2.4.1 VisualSfM: A freeware GUI application for 3D reconstruction using structure from motion (SfM). It includes improvements by the integration of both Multicore Bundle Adjustment and SIFT on the graphics processing unit (SiftGPU). Dense reconstruction can be performed through VisualSFM using PMVS/CMVS (Patch or Cluster based Multi View Stereo Software) (Kersten, 2012). During the first survey, the 218 photos taken above study area A were processed by this program with the default settings. The result is shown in Fig. 3. With this program, it is not possible to edit the high resolution color pointcloud available to Meshlab. Due to the poor results on the sparse canopy regions it was not applicable in the first survey. During the second survey, the used camera was capable of recording valuable details of the flight, such as the GPS coordinates of pictures. This meta data was also used in the second survey, that resulted in higher quality 3D models, even in the sparse regions.

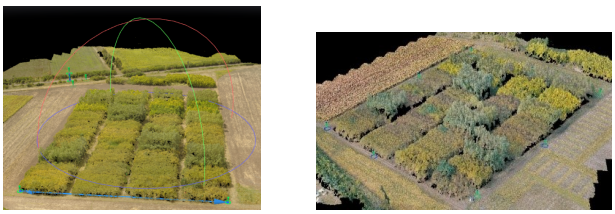


(a) 3D model of the first survey (b) 3D model of the second survey

Figure 3. The 3D models of study area A produced by VisualSfM during the two surveys.

2.4.2 Pix4D: This software proved an easy to use tool for DEM and orthophoto creation. It allows users to define orientation and scale constraints if no GPS data is accessible. This is essential when the created orthophoto and 3D model is used for calculations. In Fig. 4, the dense 3D point cloud is shown created by the software during the first survey. It gave accurate estimations even where the canopy was not continuous. All images of study area A and B were used for the 3D reconstruction process.

The model was created with the basic settings of pix4D (minimum 3 tie points for pointcloud generation, medium resolution mesh generation, etc.). Since no GPS data was associated with the photos in the first survey, we predefined one scale constraint and one orientation constraint so that the orientation and object sizes would reflect the reality. During the second survey on study area A, as a result of the photogrammetric processing, the average field resolution was 1.6 cm/pixel in the case of the RGB 3D model. The average key point per image was 75065. The average RMS error was 0.013 m. The finished 3D point cloud consists of over 19 million points, 472 points/m^2 in average. As for study area B, the average field resolution was 2.3 cm/pixel in the case of the RGB 3D model. The average key point per image was 72423. The average RMS error was 0.013 m. The finished 3D point cloud consists of over 34 million points, 2614 points/m^2 in average.



(a) Dense point cloud of the first survey (b) Dense point cloud of the second survey

Figure 4. The 3D models of study area A produced by Pix4D during the two surveys.

The photogrammetric processing of the images taken by multispectral camera was done only on study area B. The average field resolution was 7.7 cm/pixel. The average RMS error was 0.023 m. The finished 3D point cloud consists of over 412,000 points, 2.21 points/m^2 in average (see Fig. 5). Unfortunately, the photos taken by the multispectral camera were not usable in the case of study area A.

Using Pix4Dmapper photogrammetry software, the DEMs were also created during both surveys and for both study areas. During the second survey, applying the Pix4Dmapper, we also created the NDVI map of study area B. The resolution of the map is 6.9 cm/pixel and the average NDVI is 0.78. The lowest value was 0.28, and the highest one was 0.95 (see Fig. 5).

In the first survey, we could obtain an orthophoto with resolution of 8659x13378 pixels of study area A, and another one with resolution of 10145x9950 pixels of study area B. In the second survey, using a better camera the sizes of the images were 10150x8271 pixels and 10 916x15419 pixels respectively.

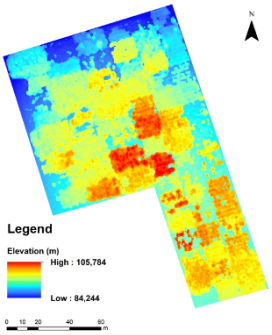
3. PREPROCESSING METHODS

When analyzing the orthophotos, it was clear that the intensity values of the leaves of the same tree species and hybrids can vary considerably. Furthermore, the canopy of trees always have small gaps between leaves, branches and crowns. To eliminate the differences, two types of blur filters were used: Gauss filter and Median filter. The last one is a nonlinear filter being used frequently to remove the “salt and pepper” image noise while preserving edges. The effect of Gaussian smoothing is also to blur the image. The degree of smoothing is determined by the standard deviation of the Gaussian and its outputs a ”weighted average” of each pixel’s neighborhood (Fisher, Wolfart, 2014). Both methods were used with different kernels.

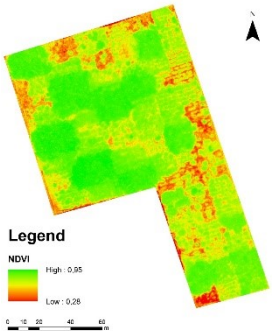
4. SEGMENTATION METHODS

The difficulty of the segmentation of species arises from the presence of hybrids. They have similar characteristics in height and color; therefore, they are barely distinguishable even to the human eye. There are a vast number of segmentation methods already in the literature. We aimed at selecting the ones that are based on color rather than edges or shapes present in images. We used the most widespread segmentation methods in the field of agriculture such as Multi-resolution segmentation provided by the eCognition software and the Mean-shift

(a) Dense point cloud of study area B



(b) NDVI map



(c) DEM

Figure 5. The 3D model and NDVI map of study area B produced by Pix4D during the second survey.

and Statistical Region Merging segmentation algorithms implemented in Matlab.

4.1 Multi-resolution segmentation (MRS)

We investigated the eCognition's hierarchical, multiresolution (MRS) algorithm. This algorithm combines pixels or objects, so it is based on region growing. Also this is an optimization method that minimizes the average heterogeneity with a given number of objects and maximizes the homogeneity of the object. It merges objects that best fit each other. The steps of the algorithm:

1. Each pixel is an independent object. These are combined in several steps into larger objects until they reach a certain homogeneity threshold. This threshold is derived from spec-tral and formal homogeneity values that can be specified in the parameter.
2. Find the best matched neighbor for each core object that is created in the first step.
3. If the best fit is not mutual, the object in the comparison will be the next object tested.
4. If the best fit is mutual, the two objects are merged.
5. In each iteration, each object is tested once.

The following parameters can be set (Hamilton, Fox, 2007):

- *Layer weights*: selection and weighting of the layers we want to apply during segmentation.
- *Scale parameter (r)*: maximum allowed heterogeneity within an object.
- *Shape (s)*: the degree of spectral and geometric homogeneity (color = 1 - shape).
- *Compactness (c)*: compactness of objects.

By changing these parameters and the input layers the size and shape of image objects are almost endlessly modifiable. The ability to perform other types of segmentation such as conditional or classification-based segmentation makes limitless modifications to the results of multi-resolution segmentation possible. Sadly, it is a semi-automatic approach, there is no generally applicable formula for assigning layer weights, setting the parameters, and implementing segmentation - ultimately, trial and error and experience are the best guides (Kersten, 2012).

4.2 Statistical Region Merging (SRM)

The Statistical Region Merging (SRM) Segmentation algorithm proposed by (Nock, Nielsen, 2004) is a time efficient method that operates as follows. It defines a real-valued function of similarity $f(p_0; p_1)$, where the p_0 and p_1 are two different points in the image. It takes each pair of points and sorts the pairs based on their similarity. In the next step, it iterates through the sorted pairs those that are not yet in the same region and merges their two regions if a predefined probabilistic function returns true. The value of the function f is based on the between-pixel local gradients, and their maximal per-channel variation.

4.3 Mean-shift

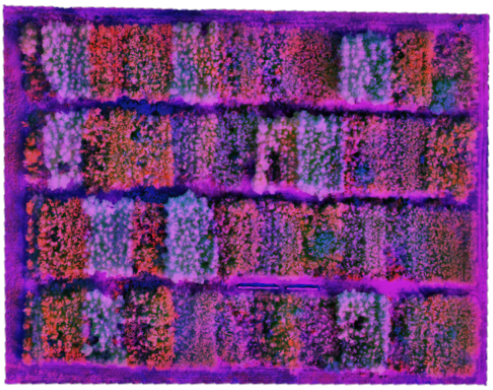
The Mean-shift Segmentation (MSS), proposed by (Fukunaga, Hostetler, 1975), is based on the assumption that the feature space is a probabilistic density function. The dense regions in the feature space correspond to local maximas. So for each data point, the algorithm performs a gradient ascent on the local estimated density until convergence. The stationary points obtained through gradient ascent represent the local maximas of the density function. All points associated with the same stationary point belong to the same cluster.

5. EXPERIMENTAL RESULTS

To be able to evaluate the results of segmentations, we created the ground truth files manually, based on the data recorded on site and the guidelines of an expert. Ground truth files have been created for both study areas in both surveys. Fig. 7 and Fig. 8 show the ground truth images of study area A and B respectively. Due to the hand drawn nature of the ground truth files, the borders of the parcels are not absolutely precise, therefore the accuracies presented in the following paragraphs contain minor human errors as well. Besides, study area B does not contain completely closed canopies, parcels contain "holes" that represent the ground and there are parcels that contain more than one species, even weed. These "holes" are represented with a color other than the parcel color in the ground truth.

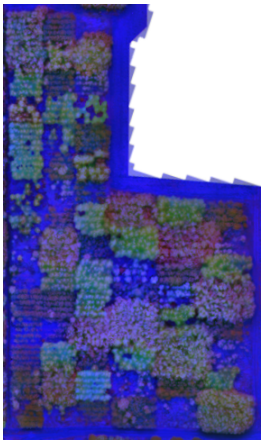
In agricultural image processing it is known that the importance of the specific color channels largely depends on the task to perform. We have followed the work of the authors of (Pap, Kiraly, 2018) and experimented with the variation of color channels with the height information encoded in the DSM image. Therefore, we created a false-color image by replacing the green channel with the height information. This was done by scaling the values of the DEM image to the range [0,255] and substituting the original green channel with it. This new false-colored image is referred to as RBD (Red Blue Digital elevation) from now on. The DEM and the RBD images are shown in Fig. **??** of study area A. During the second survey, we had access to the NIR channel and NDVI image as well in the case of study

area B. We have also used these for false-color image creation. We have experimented with combining the DEM and the NIR images with the original RGB channels.



(a) False-colored RBD image of area A

(b) False-colored RBD image of area B



(c) False-colored NBD image of area B

Figure 6. The false-colored RBD (red, green, DEM) images of area A and B, and NBD (NIR, blue, DEM) image of area B

We used the aforementioned segmentation algorithms: MRS, SRM and MSS. The goal was to find a method that is efficient but also robust in the sense that it is not strongly dependent on its input parameters. First we evaluated the eCognition's MRS segmentation algorithm that is the state of the art currently in this field of application. This is a semi-automatic process, it requires the users supervision to achieve the best results. Therefore, to reach the accuracy of this method was the goal for the other, unsupervised methods. The different segmentation methods were used with varying preprocessing methods. The segmentation methods also differ in their inputs.

5.1 Evaluation method

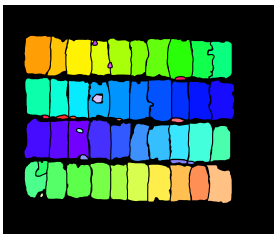
We evaluated the results of the segmentation methods by an external cluster validity index, the Sorensen-Dice similarity coefficient (D) (van Aardt, Wynne, 2004). Based on the conclusions as (Hamilton, Fox, 2007), the D is a suitable measure for evaluation in the field of biogeographic since it is less sensitive to outliers than the other coefficients. The D coefficient is calculated as follows:

$$D = \frac{2a}{(2a + b + c)} \quad (1)$$

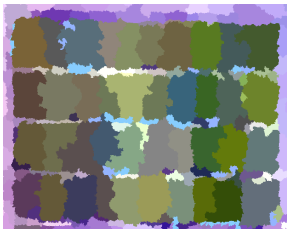
where a = the number of point pairs that belong to the same segment in the ground truth as well as in the segmentation result,
b = he number of point pairs that belong to the same segment in the ground truth, but to different ones in the segmentation result and
c = the number of point pairs that are in different segments in the ground truth, but in the same segment in the segmentation result.

6. RESULTS

The results of segmentation are presented in Table 2. In the eCognition's MRS, the user can select the RGB channels and can add the DEM and the NDVI maps of the original image as new layers. Besides, we also added the NIR band values as a layer, where it was available. The above mentioned parameters were selected empirically in our study. The image layer weights were set for RGB channels (1,5,10 and 1,10,5), for DEM (between 1 and 10), for NIR (between 1 and 10) and for NDVI (between 1 and 10). The scale parameter was tested between 100 and 240. Our shape parameter was set between 0.01 and 0.4, the compactness parameter was between 0.6 and 1.0. The original size images, and the resampled (reduced to 75%) images, did not give convincing results (68-71% accuracy). Reducing the size of the original orthophoto (and both DEM, NIR and NDVI images) to 50% led to results below 50%. The best results were obtained by performing this algorithm on the image reduced to one quarter of the original size with the layers: 1,5,10 of the RGB channels respectively, 2 for both the DEM and the NIR layer and 0 to the NDVI layer. The remaining parameters were set as follows:$r = 240, s = 0.4, c = 0.9$. After the first run of the algorithm, further hierarchical segmentation was applied manually by selecting regions that we wished to further split and the segmentation method was rerun on that region only. After performing such steps in the second survey repeatedly, we reached the highest accuracy: 76.81% for the first and 73.15% for the second area (see Fig. 7 and Fig. 8.). For the first survey, we could use neither NDVI nor NIR maps as layers, even so we reached: 74.85% for study area A and 72.55% for study area B.



(a) Ground truth

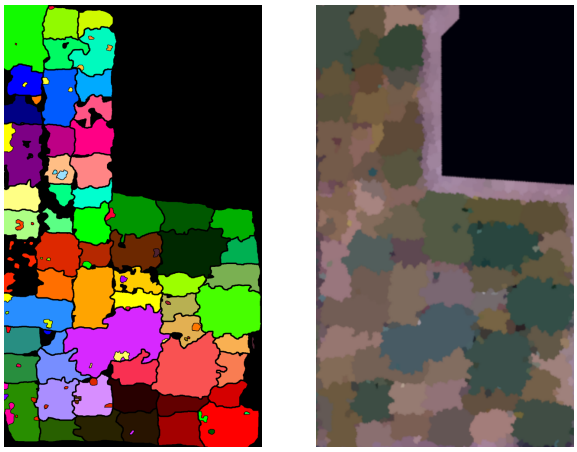


(b) Best segmentation

Figure 7. The hand crafted ground truth image and the best segmentation result on study area A

We used Matlab implementations of two segmentation algorithms: the MSS and SRM. In both cases, we resampled the original image to 32 cm/pixel spatial resolution. Furthermore, we experimented with the Median filter as a preprocessing step and tested it with kernel sizes varied between 10 and 25. We also ran Gauss smoothing filter with kernel sizes varied in the range [0-10].

We have experimented with the creation of false-colored images. We created RBD images, where we replaced the green



(a) Ground truth (b) Best segmentation

Figure 8. The hand crafted ground truth image and the best segmentation result on study area B

Study area A		
	Survey 1.	Survey 2.
eCognition (MRS)	74.85%	**76.81%**
MSS on RGB	64.15%	66.77%
MSS on RBD	66.12%	58.66%
SRM on RGB	60.03%	62.67%
SRM on RBD	61.96%	58.17%
Study area B		
	Survey 1.	Survey 2.
eCognition (MRS)	74.85%	73.15%
MSS on RGB	60.85%	**80.70%**
MSS on RBD	61.92%	72.57%
MSS on NBD	-	72.04%
SRM on RGB	58.45%	62.45%
SRM on RBD	59.45%	63.27%
SRM on NBD	-	58.61%

Table 2. Result of segmentations. Except for study area B in the 2. survey, where also the NIR channel was used, for all other cases the RGB images were used together with the DEM

channel of the RGB image with the height information encoded in the DEM image. Where the NIR band was also available, we further experimented with replacing the red channel of the RGB image with the NIR channel, and the green with the high information, thus, creating a NBD image.

The Matlab implementation of the MSS only required one parameter: the spatial radius r_s. The rest of the settings of the algorithm are calculated from this parameter. The r_s parameter was tested with the values 0.05, 0.06, ..., 0.1. The application of preprocessing methods improves the segmentation accuracy notably regardless of the type of smoothing operator. The accuracies in the case of study area A, (displayed in Table 2) were reached on images filtered by a Gauss filter of kernel size 3 and the r_s set to 0.07. In the second survey, however, using the RBD image did not improve the segmentation performance. In the case of study area B, the MSS algorithm reached the highest accuracy of 80.70% with the r_s parameter set to 0.05.

We also used the Matlab implementation of the SRM algorithm provided by (Mohan et al., 2017). The Median and Gauss smoothing filters were also tested as preprocessing steps before the application of the SRM. The algorithm takes one argument, the scale parameter Q that defines the sizes of expected regions relative to the size of the original image, by choosing the value of Q for smaller results in larger segments, while choosing it for greater results in small segments. We selected the values of Q from the range [100, 3000] on a logarithmic basis as it was proposed by (Mohan et al., 2017). In the first survey, the best results were reached on the RBD image smoothed by a Median filter with kernel size 12, where the value of Q was set to 500. In the second survey, setting the Q to 3000 gave the best results.

7. CONCLUSIONS

The aim of the paper was to investigate the applicability of a low-cost and an intermediate cost UAV in the field of tree species segmentation. We aimed to use methods that run on an image that contains several hybrids of the same tree species, a task which is hard even for the human eye. For the process of segmentation we used orthophotos that stem from the reconstructed 3D models of the two studied areas. The higher quality camera and drone resulted in more detailed, higher resolution 3D models and orthophotos. These were also georeferenced, that is a great advantage in the 3D model and orthophoto construction. Where no GPS coordinates are registered for the images, it is hard to find reliable and accurate 3D reconstruction tools to work with.

The eCognition's MRS reached 76.81% accuracy on the orthophoto based on the images taken over study area A by the UAV equipped with the better quality RGB camera. It reached 74.85% accuracy on the orthophoto produced by the UAV equipped with the lower quality RGB camera. To achieve more quality we used the NIR map as a layer although the NDVI map did not result in greater accuracy. It has been shown in the literature, that the MRS is capable of reaching even higher accuracy e.g. in (Kavzoglu, Tonbul, 2018). However, in such studies the aim is not to segment plant species apart, but rather to separate vegetation from man-made objects, such that concrete, asphalt road, roof, etc.

As it can be seen in Table 2., the use of the higher-cost UAV (survey 2.) in the case of study area A, did not produce significantly better results in segmentation. It is visible from the results, that the inclusion of the DEM information in the segmentation process in the for of RBD images, can make up for the lower quality of the images.

The fact that the higher quality images did not give us remarkably higher accuracies in segmentation is also due to the fact, that we have resampled the original orthophotos to their fraction in size. Moreover, in some cases, we also applied further smoothing filters on them. When the goal is to segment an image based on color information, small details of the images can be distractive. The higher resolution of images would mean greater advantage, when the task is to monitor plant health and discover infections. In the future, we also plan to encorporate more information from the detailed images e.g. extract textural features from them for the aid of segmentation.

ACKNOWLEDGEMENTS

The research was supported by the grant EFOP-3.6.1-16-2016-00001 "Complex improvement of research capacities and services at Eszterházy Károly University".

REFERENCES

Cunliffe, Brazier, A., 2016. Ultra-Fine Grain Landscape-Scale Quantification of Dryland Vegetation Structure with Drone-Acquired Structure-from-Motion Photogrammetry. *Remote Sensing of Environment*, 183(1), 129–143.

Fisher, Perkins, W., Wolfart, 2014. Image Processing Learning Resources. *Available online: https://homepages.inf.ed.ac.uk/rbf/HIPR2/hipr_top.htm (accessed on 10 September 2017).*

Fukunaga, K., Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1), 32–40.

Gatziolis, D., Lienard, J. F., Vogs, A., Strigul, N. S., 2015. 3D tree dimensionality assessment using photogrammetry and small unmanned aerial vehicles. *PloS one*, 10(9), e0137765.

Gaulton, R., Taylor, J., Watkins, N., 2015. Unmanned Aerial Vehicles for Pre-Harvest Biomass Estimation in Willow (Salix spp.) Coppice Plantations. *ISPRS Geospatial Week*.

Hamilton, Megown, M., Fox, 2007. Guide to automated stand delineation using image segmentation. *US Department of Agriculture, Forest Service, Remote Sensing Applications Center, Salt Lake City, Utah*.

Kavzoglu, T., Tonbul, H., 2018. An experimental comparison of multi-resolution segmentation, SLIC and K-means clustering for object-based classification of VHR imagery. *International journal of remote sensing*, 39(18), 6020–6036.

Kersten, L., 2012. Image-Based Low-Cost Systems for Automatic 3D Recording and Modelling of Archaeological Finds and Objects. *Int. Conference on Cultural Heritage, M. Ioannides et al. (Eds.), Lecture Notes in Computer Science (LNCS)*, 7616, 1-10.

Messinger, Asner, S., 2016. Rapid Assessments of Amazon Forest Structure and Biomass Using Small Unmanned Aerial Systems. *Forests*, 615.

Mohan, M., Silva, C. A., Klauberg, C., Jat, P., Catts, G., Cardil, A., Hudak, A. T., Dia, M., 2017. Individual Tree Detection from Unmanned Aerial Vehicle (UAV) Derived Canopy Height Model in an Open Canopy Mixed Conifer Forest. *Forests*, 8(9), 340.

Nock, R., Nielsen, F., 2004. Statistical region merging. *IEEE Transactions on pattern analysis and machine intelligence*, 26(11), 1452–1458.

Pap, M., Kiraly, S., 2018. Comparison of segmentation methods on images of energy plants obtained by uavs. *2018 IEEE International Conference on Future IoT Technologies (Future IoT)*, IEEE, 1–8.

Snavely, N., Seitz, S. M., Szeliski, R., 2008. Modeling the world from internet photo collections. *International journal of computer vision*, 80(2), 189–210.

St-Onge, Jumelet, C., Véga, 2004. Measuring Individual Tree Height Using a Combination of Stereophotogrammetry and Lidar. *Canadian Journal of Forest Research*, 34(1), 2122–2130.

van Aardt, J. A., Wynne, R. H., 2004. A multi-resolution approach to forest segmentation as a precursor to estimation of volume and biomass by species. *Proceedings of the American Society for Photogrammetric Engineering and Remote Sensing Annual Conference*, 24–28.

Watts, Ambrosia, H., 2012. Unmanned Aircraft Systems in Remote Sensing and Scientific Research: Classification and Considerations of Use. *Forests*, 4(6), 1671-1692.

Westoby, Brasington, G. H., Reynolds, 2012. "Structure-from-Motion" Photogrammetry: A Low-Cost, Effective Tool for Geoscience Applications. *Geomorphology*, 179, 300-314.