RESEARCH ARTICLE

An Integrated Object and Machine Learning Approach for Tree Canopy Extraction from UAV Datasets

Abhishek Adhikari[1] · Minakshi Kumar[1] · Shefali Agrawal[2] · Raghavendra S[1]

Abstract

Unmanned aerial vehicles (UAV) have emerged as new platforms for acquiring ultra-high resolution images, which are challenging for extraction of features using conventional image processing approaches. Tree canopies are required to be constantly monitored for better planning and management. UAV is currently one way to survey canopies over a large area for precisely estimating their geometry. Conventional segmentation techniques are extensively used for image feature extraction. However, they lack in accuracy and require high computational processing when used for ultra-high resolution UAV datasets. These issues can be handled by superpixel segmentation algorithms which have good boundary adherence and are computationally efficient. Simple linear iterative clustering (SLIC) is a subset of superpixel segmentation technique which uses minimum tuning parameters making it most efficient. As the random forest is known for handling multiple attributes and robustness, it can be used for classifying and extracting features from segmented image generated using SLIC. The present study mainly focuses on the automation for extraction of tree canopies along with their object-based attributes from the UAV dataset. The data acquisition was carried out using Trimble UX5 fixed-wing UAV which was further orthorectified at a spatial resolution of 13 cm. The ortho-image was further segmented using SLIC algorithm. Canopy segments are then identified and classified using random forest, which is then merged into trees objects on the basis of their proximity. Accuracy assessment was then carried out for extracted tree canopies and was found that the aforesaid approach could achieve 93% similarity index. The current study highlights the potential of using SLIC segmentation and random forest classification method for tree canopy extraction from the ultra-high resolution ortho-image derived from UAV platforms.

Keywords Tree canopy · UAV · Random forest · SLIC · Machine learning

Introduction

Trees play a vital role in sustaining the environment and one's state economy. They are major contributors in the climate amelioration, soil preservation, managing water cycle, nursing flora and fauna and survival of humans. These life-forms have always been center for studies with respect to their role in the ecology, culture and economy. To properly manage these resources, a proper evaluation of their qualitative and quantitative properties is required. There are two broadways to collect the data from the site: field survey and remote survey (Lawley, Lewis, Clarke, & Ostendorf 2016). Most of the studies on the tree are on-site field surveys (Seidel, Fleck, Leuschner, & Hammett 2011) which are time-consuming and labor-intensive. The alternative is remote sensing (RS) surveys where data are gathered using sensors fitted on terrestrial or aerial platforms (Kim, Madden, & Warner 2009). The advantage over on-site method of remotely gathering information is that they cover larger areas for which they are more economical and efficient in order to retrieve the physical aspect of the vegetation. There are three major ways of RS aerial surveys: UAV, aerial or satellite platform-based.

In recent years, unmanned aerial vehicles (UAV)-based remote sensing (UAV-RS) are being widely using for

✉ Abhishek Adhikari
  abhishek20111@gmail.com

1 Photogrammetry and Remote Sensing Department, IIRS-ISRO, Dehradun, India

2 Geospatial Technology and Outreach Programme Group, IIRS-ISRO, Dehradun, India

collecting information for various applications (Crommelinck et al. 2016). UAV-RS being able to provide geo-referenced ultra-high resolution imagery in a flexible and cost-effective way. This opens a revolutionary window to observe the surroundings with the whole new perspective (Colomina et al. 2014; Getzin, Wiegand, & Schöning 2012). UAV-RS for tree parameters estimation, at ultra-high resolution, provides users more precision and reduces the chances of errors due to poor resolution (Kaneko & Nohara 2014). At ultra-high resolution, the tree features can be distinguished from the background by their physical characteristics and patterns which can be interpreted using descriptor algorithms. The automation of the process can save time and reduce human biased errors. The image processing and classification for the UAV-RS imagery will provide the necessary algorithms to achieve full automation. There are numerous ways to delineate the objects using various techniques (Crommelinck et al. 2016; Feng, Liu, & Gong 2015) but are broadly classified into pixel-based and object-based approach. Pixel-based image processing works for low-resolution imagery, but for high resolution, the object-based approaches are preferred (Blaschke 2010). Object-based image analysis (OBIA) provides shape and size along with statistical radiometric values and hence fits the purpose of this study.

The OBIA approach is to form a more meaningful collection of pixels called objects from the whole image. This provides more information with respect to shape, size, compactness, association and statistical radiometric values of the objects. The pixels of an image, when divided into smaller dissimilar regions on the basis of spectral or spatial attributes is termed as image segmentation. Modern segmentation techniques can be classified into edge-based and region-based methods with their own set of advantages and disadvantages. In the current study, the segmentation technique that will provide fast computation cost over large datasets and requires minimum parameters tuning is required. Superpixel segmentation (Ren & Malik 2003) is a technique where the pixels are clustered into non-overlapping segments which provides a way to efficiently process over large datasets as the computation utilizes iterative local boundaries adjustments to get desired segments. In superpixels segmentation techniques, simple linear iterative clustering (SLIC) algorithm (Achanta et al. 2012) is one of the gradient ascent methods that provides a way to create superpixels with only two parameters provides a promising approach. State-of-the-art superpixels are compared in (Achanta et al. 2012; Stutz, Hermans, & Leibe 2018) in terms of boundary recall (evaluation with respect to adherence to boundaries), under-segmentation error and stability. SLIC overall was observed as the best performer overall in terms of speed, boundary adherence and under-segmentation error. A variation of SLIC SLIC-

zero (SLICO) (Achanta et al. 2012) which needs to be optimized for only one parameter (scale) is used in current work. In order to isolate canopy superpixels from the backgrounds, a classification rule that can handle multiple attributes and is robust with outcomes is needed. Random forest (RF) (Breiman, 2001) is one such assuring machine learning algorithm. The random forest algorithm is an ensemble-based machine learning algorithm. It consists of many independent decision tree classifiers which are trained on subsets of training samples; overall results of each tree are aggregated up to summarize the result. RF can handle large data with multiple input features with low latency and provide a good-fit optimal model. The RF model is fast and simple to develop and train for classification for various applications.

Literature Review

The use of UAV-RS for forest application is gaining popularity in recent years. This includes the estimation of tree physical parameters (Panagiotidis, Abdollahnejad, Surový, & Chiteculo 2017) and individual tree crown delineation (Recio, Hermosilla, A. Ruiz, & Palomar 2013). Earlier studies used the digital elevation model (DEM) derived canopy height model (CHM) (Panagiotidis et al. 2017). They were limited only to the availability of DEM, and user pre-knowledge of that region is a canopy. One of the studies used marker-controlled watershed segmentation algorithm to delineate tree canopy (Huang, Li, & Chen 2018), but this methodology required two parameters, internal and external markers, to be defined and hence lacks generalization. Furthermore, watershed segmentation subjectivity to noise sensitivity is another problem.

In a previous study (Feng et al. 2015), the authors analyzed the UAV imagery for urban vegetation mapping using pixel-based random forest and using textural-based parameters. It was observed that ultra-high spatial resolution provided enough details to classify urban vegetation from the backgrounds with high accuracy. This paper also investigated the use of OBIA but criticized the conventional OBIA approach for their complexity despite being able to give good classification results. In another study, where UAV-RS was used to study mangrove forest canopy using SLIC (Zimudzi, Sanders, Rollings, & Omlin 2018), it was observed that SLIC method provided good boundary adherence and minimum parameters optimization but alone does not produce actual objects. The over-segmented image requires a secondary clustering method to create meaningful objects. Thus, using SLIC-based algorithm simplifies OBIA and RF with textural descriptors and provides a good machine learning-based clustering approach to classify the segments and create meaningful

objects. Finally, in order to merge sub-objects (segments) using the closeness of boundaries between the superpixels, the segments can be merged into meaningful objects in this case, the tree canopies.

Materials Used

Study Area

One plot (see Fig. 1) from Uttarakhand, India was included in this work; it is located in Nahar, Koti, Dehradun. It lies in UTM zone 43 N central longitude and latitude are 782,051.684 E and 3,371,015.043 N meters, respectively. This plot consists of 109 mango, *Mangifera indica*, trees. UAV model used for surveying was UX5 and the camera used was COTS (commercially off the shelf) DSLR from SONY, Model: NEX-5 T. The tools used in the study were python, R and Arcmap.

Reference Data

The reference data are manually outlined tree canopies from UAV datasets and data collected from the field. The instrument used was "Leica Disto D8." It is a laser distometer, which uses the trigonometric-based calculation to find the height and width of the tree. The ten sample canopies were randomly chosen from the dataset and their canopy width along NS and EW direction was recorded from the site.

Workflow

The figure shows the complete conceptual diagram in brief to explain the whole methodology. It compromises of three subsections (see Fig. 2):

Part 1: Generation of ortho-images from UAV images.

Part 2: Segmentation of image and parameters generation.

Part 3: Classification of the segments and merging to form objects.

In the first section, the images collected via UAV survey were processed in Agisoft Photoscan to generate geo-referenced ortho-images as the final product. The ortho-images generated were at a spatial resolution of 13 cm. In the next section, the image is segmented using a fast superpixel pixel generation algorithm (SLIC). This is followed by parameter extraction from superpixels and its textural descriptors. In the final section, the machine learning algorithm (i.e., here random forest) is used to classify superpixels into the respective classes. The random forest is here tested for three different situations, variation with respect to input parameters:

1. Pixel-based using three bands (typ-1)
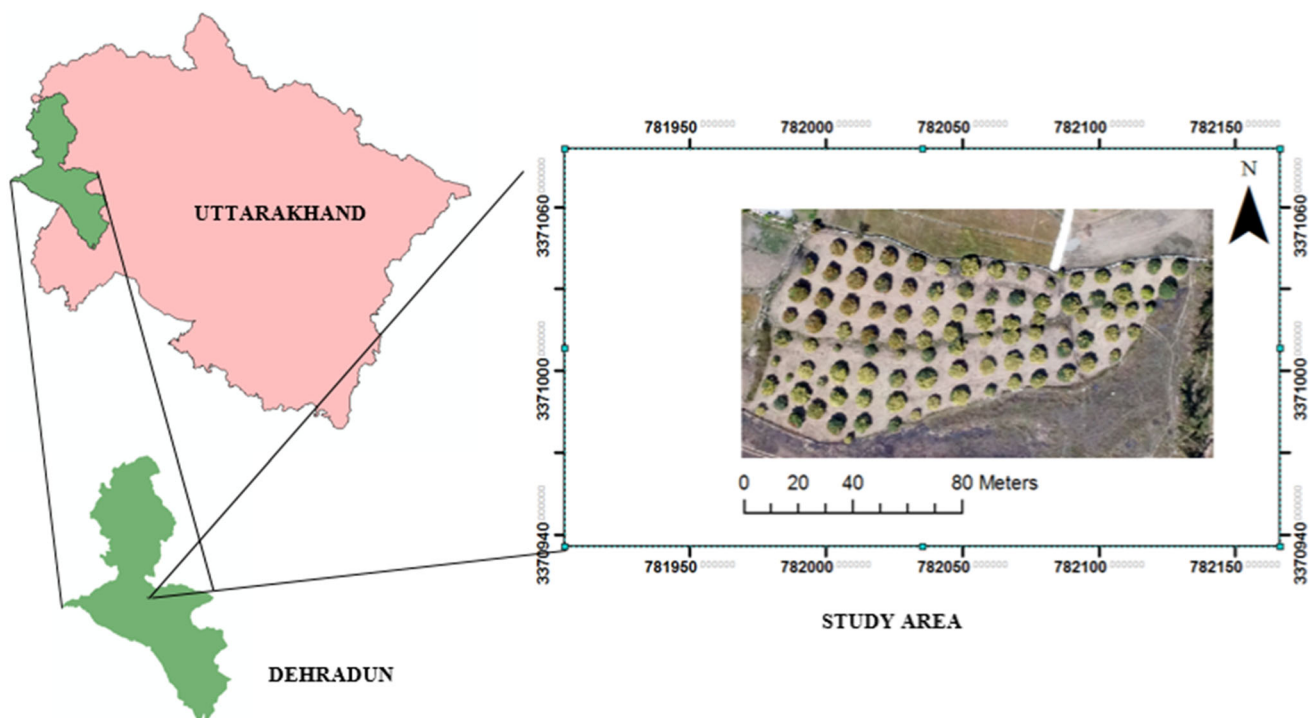2. Using only three bands (RGB + SI) (typ-2)
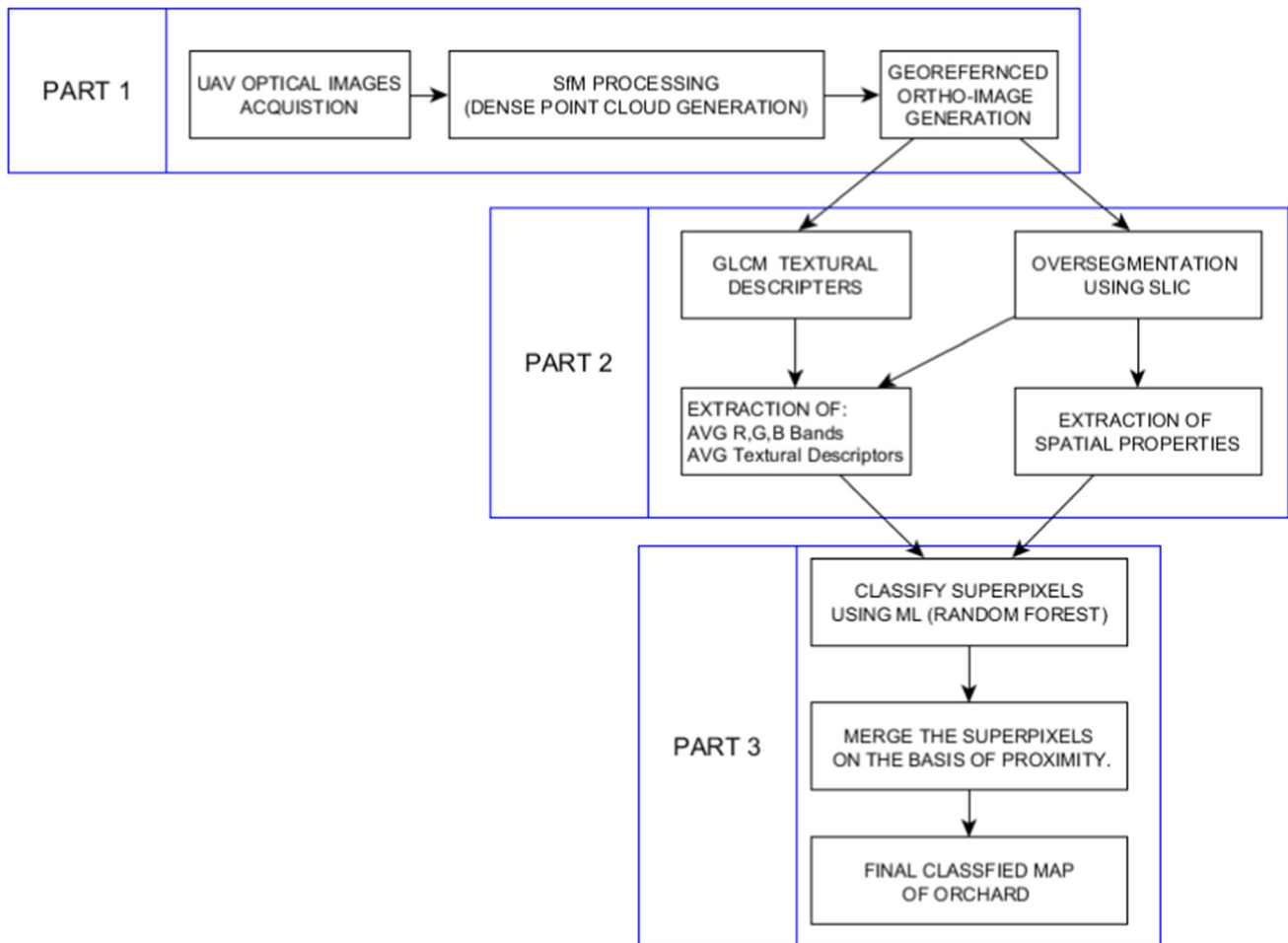

Fig. 1 Study area

Fig. 2 Methodology in brief

3. Using three bands (RGB) + seven bands (GLCM texture) + SI (typ-3. x), where x represents bin size (3, 5, 7, 9, 13, 17, 29)

After training and testing, random forest model is checked for robustness, by running model multiple times with varying training samples. The known samples were evenly distributed and divided as 30% for testing and 70% for training purpose. For typ-1, 10,000 training samples (known pixels) and three predictors for two classes (ground and canopy) were generated. For typ-2 and typ-3, 1000 samples from superpixels were used. For typ-2, four parameters (RGB + SI) were used while for typ-3, 11 parameters (RGB + SI + 7 textural parameters (mean, variance, homogeneity, contrast, dissimilarity, entropy and second moment)) were used. The two factors which check the classification quality are out-of-bag (OOB) accuracy which is calculated by correctly classifying samples out of the whole population. This factor generally overestimates model quality. The other is Cohen's kappa (kappa coefficient), which is calculated from observed and expected classes (confusion matrix) and provides more reliable validation. Finally, the tree canopy is merged on the basis of the parameters and proximity.

Results and Discussion

SLICO scale is optimized using ratio (r_{opt}) of interclass (C_{inter}) contrast and intra-class uniformity (C_{intra}) (Rosenberger, Marche, Emile, Chabrier, & Laurent 2004).

$$C_{inter} = \frac{\sum_i |S_i C_i|}{\sum_i |S_i|} \text{ where } C_i = \sum_i \frac{L_{ij}|m_i - m_j|}{m_i + m_j} \quad (1)$$

Here, C_{inter} is the sum of the contrast of all superpixels S_i. m_i and m_j are given as mean gray level of the superpixel i and j, respectively. L_{ij} is the common boundary between segments i and j. Intra-class is given as (6):

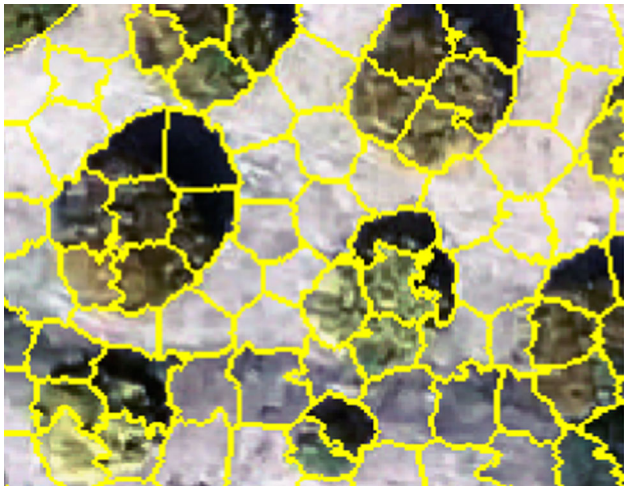$$C_{intra} = \sum_{i=1}^{k} \frac{1}{|S_i|} \left(\sum_{p \in S_i} |m_i - p| \right) \quad (2)$$

Fig. 3 SLIC segmentation of an image subset

Table 1 Confusion matrix for pixel-based RF classifier (typ-1) using testing dataset

		Predicted	
Actual		Class 1	Class 2
	Class 1	1361	62
	Class 2	80	1497

Table 2 Confusion matrix for object-based RF (typ-2) classifier

		Predicted	
Actual		Class 1	Class 2
	Class 1	149	3
	Class 2	1	147

The SLIC scale was optimized for five image subsets of different regions and different sizes. The optimized value when r_{opt} is equal to 1 for each was observed at 385,493,501, 458 and 421, respectively. The statistical mean of the five scales, 451 (approx.), was taken as final scale size for the whole image. Figure 3 represents the image subset that was segmented at the calculated scale. This value is closest for the highest common segment size (HCSS) for that particular object of interest (OI) (Fig. 4).

For the pixel-based approach (typ-1), Table 1 represents pixels-based RF classifier accuracy assessment. Here, class 1 is ground and class 2 is canopy. In this case, the best accuracy was achieved that was 95.4% (OOB-accuracy) and using confusion matrix 95.3%. Very high accuracy was achieved in typ-1 but there were gaps within canopy classified as ground. OBIA method was tested in two ways without (typ-2) and with texture (typ-3).
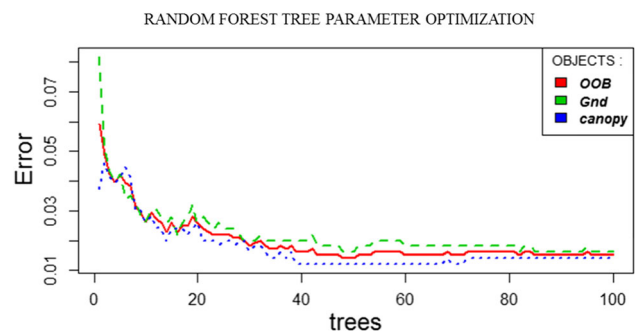

Fig. 5 Optimization of random forest n tree (number of trees) for the canopy extraction when only RGB and 7 GLCM band + SI used as an input parameter for bin size 9

OBIA-RF typ-2 classifier gave the best classification result of 98.8% (OOB accuracy) and 98.6% kappa accuracy. At 50 trees, RF results in minimum error rate hence 50 was taken as n trees optimized value for RF classifier.

Fig. 4 Final classified image (for typ3.9), highest accuracy amongst typ-3 was obtained at bin size 9

Table 3 Confusion matrix for object-based RF classifier at bin size 9 (typ-3.9)

		Predicted	
Actual		Class 1	Class 2
	Class 1	149	4
	Class 2	1	146

The typ-2 result was taken for other parameters extraction (Fig. 7). Table 2 represents the accuracy using the confusion matrix.

When typ-3 (11 parameters OBIA-RF) is applied, similar results to typ-2 were observed. Among different bin sizes, typ-3.9 gave best classification result of 98.14% (OOB accuracy) and kappa coefficient of 98.3% in all bin sizes. Figure 5 represents the optimization for the number of trees required for this process. At 50 trees, RF results in
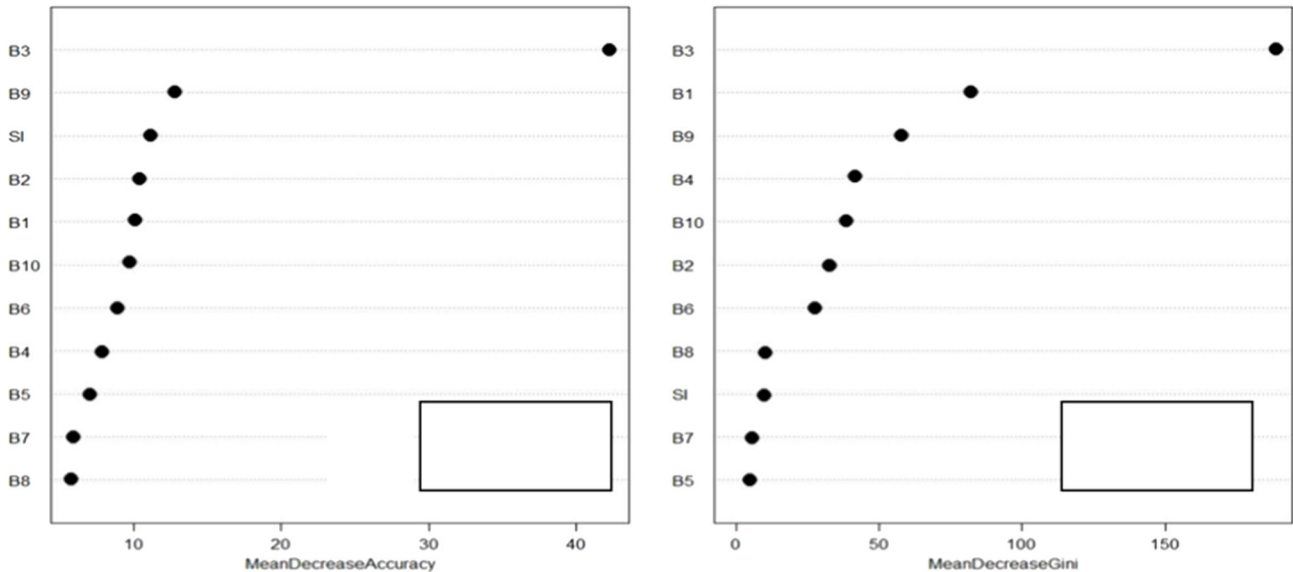
B1 = RED	B5 = VARIANCE	B9 = ENTROPY
B2 = GREEN	B6 = HOMOGENEITY	B10 = SEC. MOMENT
B3 = BLUE	B7 = CONTRAST	SI = SHAPE INDEX
B4 = MEAN	B8 = DISSIMILARITY	

Fig. 6 Mean decrease in accuracy (**a**) and mean decrease in Gini index (**b**) bin size = 9

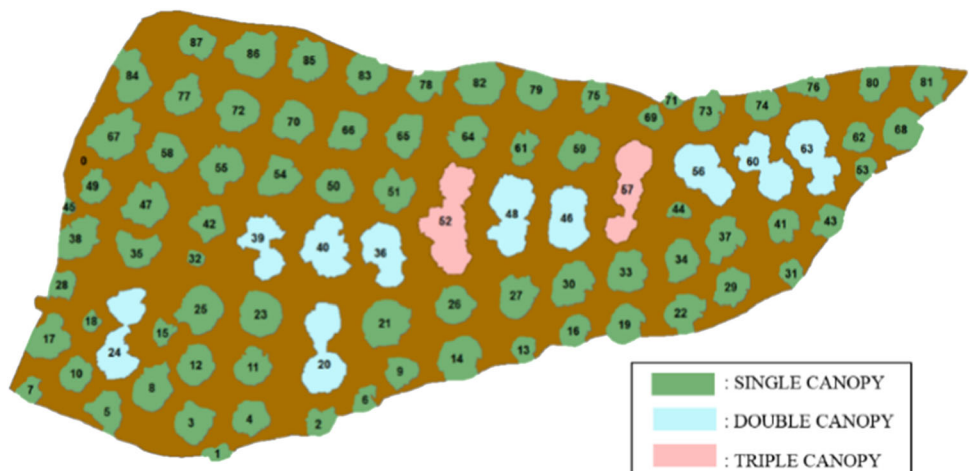**Fig. 7** Further classification into the single double and triple canopy of the tree canopy

Table 4 Area of 10 trees canopies from three sources (in sq. meters)

S No	Area (auto)	Area (manual)	Area (field)
1	29.297	28.191	28.938
2	35.336	33.802	34.628
3	29.938	29.102	29.875
4	38.799	36.929	39.147
5	43.920	41.734	44.120
6	31.737	30.229	32.120
7	28.544	28.919	28.843
8	30.871	27.362	30.876
9	7.180	5.872	7.092
10	29.527	33.794	28.440

minimum accuracy. Table 3 represents the accuracy using the confusion matrix.

Figure 6 represents the minimum decrease in accuracy (MDA) and a mean decrease in Gini index (MDGI). More the accuracy of the random forest decreases due to the removal of a single variable, the more important the variable, and therefore variables with a large mean decrease in accuracy (MDA) played a significant role in classification. Figure 6. (A) Depicts band 3 (blue) as the most significant variable. This is followed by B9 (entropy) and then shape index (SI).

Mean Gini importance measures the average gain of purity by splits in trees for a given variable. More the variable splits classes purely higher the value. Fig. 7. (B) Shows B3 (blue) as the most significant variable followed by B1 (red band) and then B9 (ENTROPY). The equation was also validated for different bin sizes versus their accuracy assessment. Overall, typ-2 and typ-3 performed better than typ-1 as in object-based noise is concealed within the object leading to more refined structures.

This classified results of typ-2 and typ-3.9 were tested against the manually using equation of similarity index: Eq. (3). Where XOR is the "exclusive or" operator. The similarity between typ-3.9 and manually drawn canopy observed is 92.9%. The similarity between typ-2 and manual is 93%. Both typ-2 and typ-3 methods gave quite close results to manual classification.

$$\text{Similarity Index (SI)} = \frac{\sum_0^x \text{XOR}(\text{Mask}_{\text{org}}, \text{Mask}_{\text{mod}})}{\text{total elements}} \quad (3)$$

where Mask_{org} (manual handmade mask) and Mask_{mod} (mask generated using RF-SLIC) for all $x \in$ pixels in region compared.

The elements are finally merged on the basis of proximity and border they share. In the result of using typ-2 classifier was used (see Fig. 7). Figure 7 represents the final classified result. In this, the image is classified into single, double and triple canopy manually and is correlated to the shape attributes like shape index. In order to compare the shapes generated, the calculated area's of selected canopies from the field, manually drawn shapes and shapes generated using SLIC-RF (typ-2) method are shown in Table 4. It was observed that the area calculated from the automatically generated canopies had an average difference of 0.354 sq. meters. And the standard deviation of 0.311 while manually generated canopies had had an average difference of 1.9 sq. meters and the standard deviation of 1.49.

Conclusion

This study investigates consumer grade optical sensor UAV for delineation of tree canopies using SLIC-RF automation process. This supports the use of OBIA and machine learning (ML) in various UAV-RS application for parameters extraction. It was also observed that superpixels' size for an OI is independent of image size at fixed spatial resolution. As r_{opt} tends to one, the best size of superpixel is achieved at which algorithm is computationally efficient and also OI is not under-segmented. This methodology also simplifies the OBIA approach by integrating SLICO-RF to create meaningful objects with minimum parameters optimization. As both typ-2 and typ-3 perform nearly the same, the RGB + SI provides a faster alternative at current spatial resolution rather than using additive textural parameters. This is also supported by the MDA and MDGI charts where the blue band is depicted as a major contributor to classification. The current methodology fails to separate overlapping canopy structures. Also, the inclusion of infrared band and DEM can provide a better insight into the delineation of canopies. The current methodology overall provides a new perspective to measure the physical aspects of tree canopy using UAV-RS.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(11), 2274–2281. https://doi.org/10.1109/TPAMI.2012.120

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing Elsevier.* https://doi.org/10.1016/j.isprsjprs.2009.06.004

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Colomina, I., Molina, P., Harwin, S., Lucieer, A., Skarlatos, D., Procopiou, E., & Eisenbeiss, H. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing,*

92(November), 79–97. https://doi.org/10.1016/j.isprsjprs.2014.02.013

Crommelinck, S., Bennett, R., Gerke, M., Nex, F., Yang, M. Y., & Vosselman, G. (2016). Review of automatic feature extraction from high-resolution optical sensor data for UAV-based cadastral mapping. *Remote Sensing, 8*(8), 689. https://doi.org/10.3390/rs8080689

Feng, Q., Liu, J., & Gong, J. (2015). UAV remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sensing, 7*(1), 1074–1094. https://doi.org/10.3390/rs70101074

Getzin, S., Wiegand, K., & Schöning, I. (2012). Assessing biodiversity in forests using very high-resolution images and unmanned aerial vehicles. *Methods in Ecology and Evolution, 3*(2), 397–404. https://doi.org/10.1111/j.2041-210X.2011.00158.x

Huang, H., Li, X., & Chen, C. (2018). Individual tree crown detection and delineation from very-high-resolution UAV images based on bias field and marker-controlled watershed segmentation algorithms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11*(7), 2253–2262. https://doi.org/10.1109/JSTARS.2018.2830410

Kaneko, K., & Nohara, S. (2014). Review of effective vegetation mapping using the uav (unmanned aerial vehicle) method. *Journal of Geographic Information System, 6*(6), 733–742. https://doi.org/10.4236/jgis.2014.66060

Kim, M., Madden, M., & Warner, T. T. (2009). Forest type mapping using object specific texture measures from multispectral ikonos imagery segmentation quality and image classification issues. *Photogrammetric Engineering & Remote Sensing, 75*(7), 819–829. https://doi.org/10.14358/PERS.75.7.819

Lawley, V., Lewis, M., Clarke, K., & Ostendorf, B. (2016). Site-based and remote sensing methods for monitoring indicators of vegetation condition: An Australian review. *Ecological Indicators, 60,* 1273–1283. https://doi.org/10.1016/J.ECOLIND.2015.03.021

Panagiotidis, D., Abdollahnejad, A., Surový, P., & Chiteculo, V. (2017). Determining tree height and crown diameter from high-resolution UAV imagery. *International Journal of Remote Sensing, 38*(8–10), 2392–2410. https://doi.org/10.1080/01431161.2016.1264028

Recio, J., Hermosilla, T., Ruiz, A., & L., & Palomar, J. . (2013). Automated extraction of tree and plot-based parameters in citrus orchards from aerial images. *Computers and Electronics in Agriculture, 90,* 24–34. https://doi.org/10.1016/j.compag.2012.10.005

Ren, X., & Malik, J. (2003). Learning a Classification Model for Segmentation. in *Proceedings of the Ninth IEEE International Conference on Computer Vision-Volume 2* (Vol. 13, p. 10). Washington, DC, USA: IEEE Computer Society. https://doi.org/https://doi.org/10.21313/hawaii/9780824839673.003.0030

Rosenberger, C., Marche, P., Emile, B., Chabrier, S., & Laurent, H. (2004). Unsupervised evaluation of image segmentation application to multi-spectral images. *Proceedings of the 7th International Conference on Pattern Recognition.* 576–579 Vol 1. https://doi.org/https://doi.org/10.1109/icpr.2004.1334206

Seidel, D., Fleck, S., Leuschner, C., & Hammett, T. (2011). Review of ground-based methods to measure the distribution of biomass in forest canopies. *Annals of Forest Science, 68*(2), 225–244. https://doi.org/10.1007/s13595-011-0040-z

Stutz, D., Hermans, A., & Leibe, B. (2018). Superpixels: An evaluation of the state-of-the-art. *Computer Vision and Image Understanding, 166,* 1–27. https://doi.org/10.1016/j.cviu.2017.03.007

Zimudzi, E., Sanders, I., Rollings, N., & Omlin, C. (2018). Segmenting mangrove ecosystems drone images using SLIC superpixels. *Geocarto International.* https://doi.org/10.1080/10106049.2018.1497093