ISPRS Journal of Photogrammetry and Remote Sensing 102 (2015) 14–27

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Training set size, scale, and features in Geographic Object-Based Image Analysis of very high resolution unmanned aerial vehicle imagery

Lei Ma [a,d], Liang Cheng [a,b,c,d,*], Manchun Li [a,b,d,e,*], Yongxue Liu [a,b,d,e], Xiaoxue Ma [d]

[a] Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University, Nanjing 210023, China
[b] Collaborative Innovation Center for the South Sea Studies, Nanjing University, Nanjing 210023, China
[c] Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University, Nanjing, China
[d] School of Geographic and Oceanographic Sciences, Nanjing University, Nanjing 210023, China
[e] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China

ARTICLE INFO

Article history:
Received 28 May 2014
Received in revised form 20 December 2014
Accepted 24 December 2014

Keywords:
GEOBIA
OBIA
Scale
Training set size
UAV
Very High Resolution (VHR)

ABSTRACT

Unmanned Aerial Vehicle (UAV) has been used increasingly for natural resource applications in recent years due to their greater availability and the miniaturization of sensors. In addition, Geographic Object-Based Image Analysis (GEOBIA) has received more attention as a novel paradigm for remote sensing earth observation data. However, GEOBIA generates some new problems compared with pixel-based methods. In this study, we developed a strategy for the semi-automatic optimization of object-based classification, which involves an area-based accuracy assessment that analyzes the relationship between scale and the training set size. We found that the Overall Accuracy (OA) increased as the training set ratio (proportion of the segmented objects used for training) increased when the Segmentation Scale Parameter (SSP) was fixed. The OA increased more slowly as the training set ratio became larger and a similar rule was obtained according to the pixel-based image analysis. The OA decreased as the SSP increased when the training set ratio was fixed. Consequently, the SSP should not be too large during classification using a small training set ratio. By contrast, a large training set ratio is required if classification is performed using a high SSP. In addition, we suggest that the optimal SSP for each class has a high positive correlation with the mean area obtained by manual interpretation, which can be summarized by a linear correlation equation. We expect that these results will be applicable to UAV imagery classification to determine the optimal SSP for each class.

1. Introduction

Geographic Object-Based Image Analysis (GEOBIA) is a systematic framework for geographic object identification, which combines pixels with the same semantic information into an object, thereby generating an integrated geographic object, before recognizing the geographic object using GIS spatial analysis or a mature classification algorithm, i.e., Neural Networks (NN), Maximum Likelihood (ML), Support Vector Machines (SVM), and Random Forests (RF). GEOBIA is also a new and evolving paradigm, which was designed specifically for high resolution remote sensing image data, in contrast to the pixel-based approach (Benz et al., 2004; Liu et al., 2006; Blaschke, 2010; Myint et al., 2011; Addink et al., 2012; Blaschke et al., 2014). Indeed, GEOBIA has become a popular alternative for land cover and land use classification (Radoux and Bogaert, 2014). Since the first international GEOBIA conference in Calgary, Canada, the unique advantages of GEOBIA have attracted the attention of scholars throughout the global field of remote sensing (Hay and Castilla, 2008; Powers et al., 2012; Arvor et al., 2013; Costa et al., 2014; Blaschke et al., 2014). GEOBIA has many advantages as a new paradigm in the diverse fields of remote sensing because it is readily combined with GIS to provide a complete vector map of land use types, which can be used directly for GIS analysis (Arvor et al., 2013). However, the pixels that belong to the same object cannot be combined into one complete object accurately due to the uncertainty of segmentation, which is a process used to partition a complex image scene into non-overlapping homogeneous regions (Witharana and Civco, 2014). This is because over-segmentation and under-segmentation

* Corresponding authors at: Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University, Nanjing 210023, China. Tel.: +86 25 8359 7359 (L. Cheng), +86 25 8359 5336 (M. Li).
E-mail addresses: lcheng@nju.edu.cn (L. Cheng), limanchun_nju@126.com (M. Li).

http://dx.doi.org/10.1016/j.isprsjprs.2014.12.026
0924-2716/

always occur due to inappropriate segmentation parameters, especially the Segmentation Scale Parameter (SSP) (Kim et al., 2011; Zhang et al., 2013). Thus, many problems are caused by segmentation, most of which are also known to affect pixel-based image analysis. These problems include the strategies used for sample selection, feature selection, accuracy assessment, and change detection, as follows. (1) Which of the classes should we label for over- or under-segmentation during sampling? (2) What are the best features for classification and what is the most appropriate feature selection method (e.g., spectral, textural, geometrical, or semantic features). (3) In GEOBIA, the planar vector layer leads to many objects being generated by segmentation; thus, should we consider these objects as single points to evaluate the accuracy, or return to evaluation at the pixel level by stacking the classified layer and reference layer? In other words, is the point-based accuracy evaluation method or the area-based method more suitable, and how does the segmentation scale affect them?

In terms of the scale problem, multiresolution segmentation (MRS) has proved to be one of the most successful image segmentation algorithms in the GEOBIA framework (Witharana and Civco, 2014). This algorithm is relatively complex and user-dependent, where the scale, shape, and compactness are the main parameters manipulated by users to control the algorithm (Witharana and Civco, 2014). The scale parameter is considered to be one of the most important variables because it controls the relative size of the image objects, which has a direct impact on the subsequent classification steps (Kim et al., 2011; Myint et al., 2011; Hussain et al., 2013; Drăguţ et al., 2014). Blaschke et al. (2014) also note that semantically significant regions are found at different scales, which makes it important to adjust the scale parameter during segmentation to obtain optimal results. However, many of the specific applications used for identification rely on trial and error to determine the optimal scale parameter based on the experience of the operators (Laliberte and Rango, 2009; Stefanski et al., 2013; Ma et al., 2014; Witharana and Civco, 2014). Clearly, this approach is not desirable because it is user-dependent (Johnson and Xie, 2011). Thus, many methods have been proposed for determining the scale parameter (Drăguţ et al., 2010, 2014; Johnson and Xie, 2011). However, most of these proposed methods are based on specific imagery and none considers the actual cover characteristics when determining the optimal SSP. To implement a multiscale hierarchical classification method, we usually need to set different segmentation scale parameters from top to bottom, thereby ensuring the extraction of objects with different sizes (Kim et al., 2011; Duro et al., 2012). This means that the optimal segmentation scale is different for various ground objects (e.g., cropland, buildings, water bodies, and woodland). Therefore, we tried to consider the actual size of the ground objects to determine their optimal SSP in the present study. We also aimed to elucidate the specific relationship between the optimal SSP and the characteristics of actual objects.

Multiscale GEOBIA can generate dozens and sometimes hundreds of variables for classifying imagery (Duro et al., 2012). In particular, for UAV VHR imagery, the number of features in each object can exceed 200 at each scale with eCognition software, which increases at finer scales due to the soaring number of segmented objects. Analyzing these features can be more computationally intensive than the analysis of photos obtained from piloted aircraft or satellites (Laliberte and Rango, 2009). The high number of features also complicates the construction of a classifier and it leads to the curse of dimensionality or Hughes phenomenon (Pal and Foody, 2010). Therefore, feature selection is an important step when improving the accuracy and efficiency of classification. Two goals of feature selection are obtaining a more thorough understanding of the underlying processes that influence the data and identifying discriminative and useful features for classification and prediction (Guyon and Elisseeff, 2003). Feature analysis based on pixels is performed more frequently compared with GEOBIA (Novack et al., 2011). Several feature reduction techniques are also used frequently in remote sensing, including InfoGain (Novack et al., 2011), Relief-F (Novack et al., 2011), RF algorithms (Pal and Foody, 2010), Correlation-based Feature Selection (CFS) (Hall et al., 2009), and principal components analysis (Pal and Foody, 2010). There have been no previous evaluations of the performance of these methods with GEOBIA, except a comparison of three unrepresentative feature selection methods reported by Laliberte et al. (2012). In the present study, we did not focus on the scale of the feature analysis. Instead, we simply employed the filtering algorithm CFS, which can select a feature subset using a correlation-based heuristic evaluation function (Hall et al., 2009), to obtain a subset of the best selected features (Pal and Foody, 2010).

GEOBIA must also overcome similar challenges to the traditional pixel-based approaches, such as the training set size and its completeness, where the image objects are initially extracted from an image (Pal and Mather, 2003; Congalton and Green, 2009; Hussain et al., 2013). It is essential that the number of classes is adequate for describing the land cover of the study area and the training data must provide a representative description of each class (Pal and Mather, 2003). For example, an important requirement of the ML classifier is that the number of pixels included in the training dataset for each class should be at least 10–30 times the number of features. Many previous studies have indicated that the size of the training dataset has a substantial effect on the classification accuracy (Pal and Mather, 2003; Foody et al., 2006). In traditional pixel-based classification analysis, there have been many analyses of the training set size (Van Niel et al., 2005; Foody et al., 2006; Rodriguez-Galiano et al., 2012), but few studies have addressed this issue for GEOBIA. For example, Zhen et al. (2013) investigated the effect of the training set size on the classification accuracy and the accuracy estimates obtained from the validation data, where the training and validation data were obtained from several selection schemes using WorldView-2 data. However, a key consideration is the deficiency of accuracy evaluation methods based on points, rather than area-based or polygon-based methods, as recommended recently, which may be a more reasonable evaluation method for segmented objects (Whiteside et al., 2014; Radoux and Bogaert, 2014). The accuracy assessment sample units may include single pixels, blocks, and polygons (Stehman and Wickham, 2011), but the accuracy assessment sample units should be polygons if the polygon map is created by manual interpretation, or using image segmentation and object-based classification algorithms (Congalton and Green, 2009). However, Zhen et al. (2013) failed to consider the effect of the training set size on the accuracy of classification using different segmentation scales. Thus, to make our results more robust, we analyzed the effect of the training set size on the classification accuracy, before using the classification accuracy to evaluate the scale. We also determined how the training set size affects the accuracy at different segmentation scales in GEOBIA.

In addition, civilian applications of Unmanned Aerial Vehicle (UAV) have increased considerably in recent years due to their great availability and because of the miniaturization of sensors, GPS, inertial measurement units, and other hardware (Zhou et al., 2009). UAV have been combined with remote sensing technology to acquire spatial data related to land cover, resources, and the environment for use in remote sensing data modeling and analysis processes (Cheng et al., 2008, 2012, 2014; Zhang and Kovacs, 2012; Ma et al., 2013a; Gomez-Candon et al., 2014). The high frequency and Very High Resolution (VHR) images obtained using UAV means that UAV have received increasing attention from researchers and manufacturers (Laliberte and Rango, 2009; Jaakkola et al., 2010; Kim et al., 2013; Ma et al., 2014; Lucieer et al., 2014). Thus, the

application range of UAV has expanded to include forest resource management and monitoring (Dunford et al., 2009), vegetation and river monitoring (Sugiura et al., 2005), disaster management, especially earthquake monitoring (Dong and Shan, 2013), and precision agriculture (Zhang and Kovacs, 2012). However, individual UAV images generally cover a relatively small area with very high resolution, since UAVs usually fly with at a relatively low elevation (Ma et al., 2013b). Due to the challenges in the data preprocessing, such as triangulation, orthorectification and mosaicking (Ma et al., 2013b), it may be preferable to classify such UAV images in small scenes, rather than as a large mosaic. Moreover, in classifying UAV images individually, the analyst maybe face with issues regarding the number of training samples to select and the distinct scale effect. This is because with small individual images, the total population of segments from which to draw samples is finite. This problem may be particularly acute, as the segmentation scale becomes greater, and the number of segments decreases.

To address these problems, we propose a technical framework for UAV classification using GEOBIA, where we consider the strategies used for optimal sample selection, efficient feature selection, and robust accuracy assessment, as follows. First, manual interpretation is applied to UAV imagery to acquire prior knowledge of the imagery, before using a stratified random sampling strategy to assess the effect of the training set size. Second, a RF classifier is used to classify the UAV imagery and an area-based accuracy evaluation method is employed to assess the sensitivity to the training set size and scale variations. We hope that the derived laws or theory may be useful for GEOBIA, by providing scientific support for researchers and users of other VHR imagery types. Our results are incorporated into a workable solution for classification using UAV.

2. Study area and data acquisition

The study area selected was Deyang, which is located in the northeast hilly area of the Chengdu plain in Sichuan Province, China, and this area includes cropland, woodland, and rural buildings (Fig. 1). We used a fixed wing UAV with a digital camera,

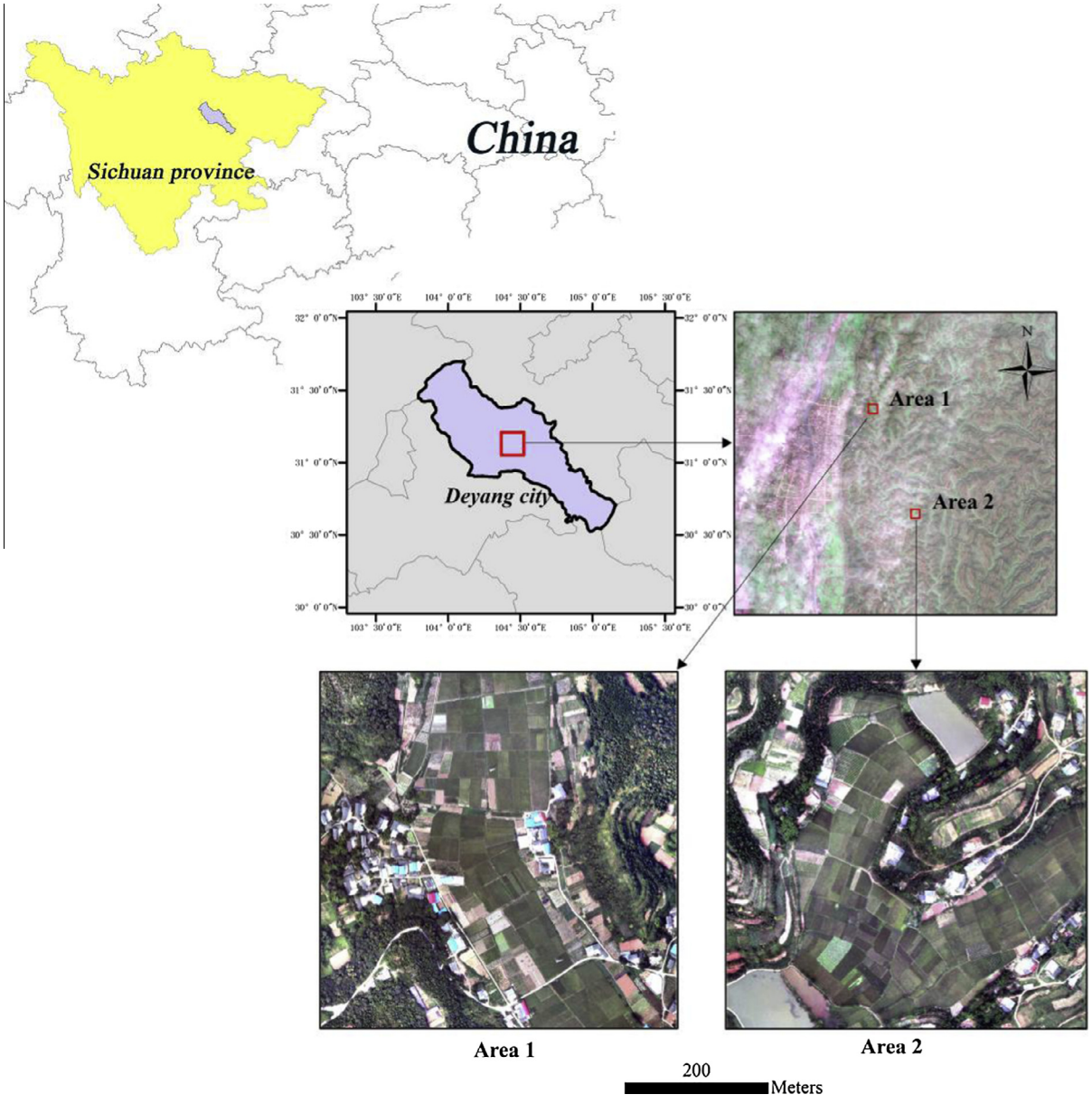

Fig. 1. Study areas.

Canon EOS 5D Mark II, to acquire images with an 80% forward lap and a 60% side lap. The images were acquired at a height of 750 m, August 2011. The resolution of the camera was 5616 × 3744 pixels, which generated an image size of 1123 × 748 m, with a pixel resolution of 0.2 m. The focal length of the camera was 24.5988 mm and the pixel size was 0.0064 mm. Subsequently, using the collected control points, a digital orthophoto map (DOM) with a 500 × 500 m standard sheet was produced by digital photogrammetry (Ma et al., 2013a). It should be noted that UAV data differ from satellite imagery in terms of navigation, orientation, and the remote sensing payloads (Colomina and Molina, 2014). The effects of navigation and orientation mean that the imageries always require preprocessing, such as triangulation, orthorectification, and mosaicking. Moreover, the UAV equipped with a Canon EOS 5D Mark II could only acquire RGB bands.

Both of the standard sheets were used in our study, i.e., area 1 and area 2 (Fig. 1). Study area 1 comprised a variety of land cover types, but mainly cropland (38%) with paddy fields and dry land. The remainder of study area 1 was characterized by the presence of woodland (43%), buildings (6%), bare land (5%), and roads (2%). Study area 2 comprised cropland (45%), woodland (37%), buildings (4%), bare land (4%), roads (1%), and water (5%). The low percentages of roads in both images meant that they were not considered during classification because of the low sample sizes. Thus, only cropland, woodland, buildings, bare land, and water were classified.

3. Methods

Based on the DOM, the general workflow (Fig. 2) was as follows. (1) Segment the image at multiple scales. (2) Sample using a stratified random scheme, with the same training set ratio (proportion of the segmented objects used for training) for each class. (3) Select features for each scale using CFS. (4) Classify the image using a RF classifier. (5) Evaluate the performance accuracy using an area-based method. It should be noted that the training set size was evaluated before classification. The initial assessment of the UAV images for GEOBIA was used to set the training set ratio because no specialized evaluation procedure is available at present to select a suitable training set for GEOBIA. Thus, the training set size was evaluated at each scale to ensure that the classification accuracy evaluation obtained robust results at all scales. The precision was also comparable at different scales.

3.1. Image segmentation

We used eCognition 8.7 (eCognition Software® Definiens, 2011) to segment the UAV image scenes. The eCognition developer's proprietary MRS is known to be one of the most successful segmentation algorithms and it is capable of producing meaningful image objects in many remote sensing application domains. This segmentation method is a bottom-up region merging technique, where small segments are merged into larger ones based on the heterogeneity (similarity of spectral and spatial characteristics) of adjacent objects in terms of three parameters: scale, shape, and compactness. The scale parameter defines the maximum standard of the homogeneity criteria with respect to the weighted image layers used to segment objects, where a higher value yields larger objects. The composition of the homogeneity criteria is controlled by the color and shape, where the sum of the weight coefficient for each pair is equal to 1. The shape parameter defines the textural homogeneity used to derive objects, including the smoothness and compactness (shape = smoothness + compactness) (eCognition Software® Definiens, 2011). Therefore, the actual composition of the homogeneity criteria can be customized by weighting the shape and compactness criteria.

In this study, a segmented image series was obtained using 20 different scale parameters, which started with a scale parameter of 10 and ended at 200, with increments of 10, which allowed the image to be analyzed at multiple scales. The weights of color/shape and smoothness/compactness were the same for every scale. Color/shape was set to 0.9/0.1 because we wanted the spectral information to have the most important role during segmentation. Smoothness/compactness was set to 0.5/0.5 because we did not

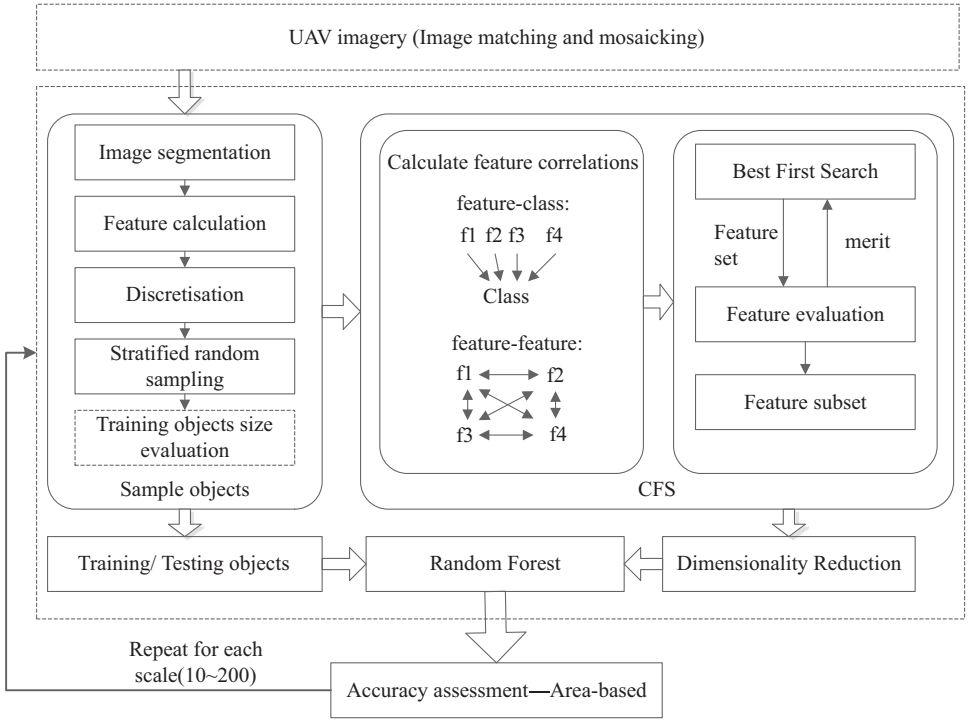

Fig. 2. Framework of the overall procedure.

want to favor compact or non-compact segments. The weights of the image layer were set to 1 for all three bands to avoid any bias.

3.2. Classification

3.2.1. Sample selection

We reviewed previous studies of accuracy assessment to facilitate the design of the sampling and labeling approach, because most accuracy assessment approaches are highly dependent on the sample selection during the classification of remote sensing images. Pixels, blocks of pixels, and polygons are all potentially viable spatial units when conducting an accuracy assessment according to the pixel-based paradigm (Stehman and Wickham, 2011). Simple random, stratified random, systematic, and cluster sampling are considered by many sampling schemes (Congalton and Green, 2009). In addition, "clusters of polygons" is a sample unit utilized by GEOBIA. According to the GEOBIA paradigm, the sample units and sample schemes determine whether the accuracy assessment results are reliable or not because of the uncertain target objects (i.e., the object size changes with the scale). Congalton and Green (2009) recommend polygons as the usual sample units for object-based classification. However, the use of polygons as sample units can cause confusion (Fig. 3), because there are several possible topological relationships between the classified objects and reference objects (Whiteside et al., 2014). Thus, a method needs to be developed for creating the sample object label (discussed in Section 4.2), where the simplest approach uses the majority class of the polygon to create the object label (Congalton and Green, 2009). In addition, simple random sampling was used to obtain reference data for training and assessment in most previous studies (Zhen et al., 2013; Puissant et al., 2014; Whiteside et al., 2014). However, this approach is not always appropriate because it tends to under-sample rare but possibly very important map categories unless the sample size is notably increased. Thus, we recommend stratified random sampling, which ensures unbiased sample selection and adequate numbers of samples in each class.

The results of the manual interpretations of UAV imagery were used as a reference layer that overlapped the segmented layer, where a polygon in the segmented layer belonged to a class if it contained more than 60% of the specific class. Subsequently, the sample sets were constructed automatically at each scale by software based on ArcEngin9.3 and C#. Thus, stratified random sampling was implemented at the sample set level, where all of the segmented layers were labeled to link with the reference layers. However, there was a problem when the sample set was used to train the rules for classification, which depended on the appropriateness of the ratio at all scales to make them comparable. Section 3.2.2 provides a quantitative analysis of the optimal training set size.

3.2.2. Evaluation of the optimal training dataset size

The number of training areas should be as large as possible to represent all of the variability in a category (Pal and Mather, 2003), but it is also necessary to design a sampling scheme that performs well in economic and time terms, which can obtain an acceptable mapping accuracy level (Lippitt et al., 2008). In this study, we employed a unified and comparable scheme to determine the training set size, which was linked to the scale (i.e., the effects of the number of objects and the average area); thus, an appropriate evaluation of the training set size was necessary before classification. We aimed to acquire a reasonable of training set ratio for scales of 10–200, which differed from previous studies. More specifically, we ensured that the variables used at all scales remained efficient when the area-based accuracy was used. Given the requirement for high-resolution image classification and accuracy evaluation, we used geographic object-based technology; thus, the minimum unit was a homogeneous object (polygon) rather than a pixel (Stehman and Wickham, 2011; Radoux and Bogaert, 2014). The optimization of the training set aimed to find a comparable variable for the training set size using a RF classifier, which was not certain to achieve the highest classification accuracy. This specific evaluation of the training set size using UAV imagery before feature selection and RF classification aimed to make our results more reliable without strong hypotheses, but this evaluation also provides guidance for selecting the training set size with GEOBIA using VHR imagery in the future. To reduce the uncertainty of the classification accuracy due to the use of different training set sizes at different scales, we employed a predictor of the overall accuracy (OA) during the assessment of the training set size for GEOBIA to determine the optimal training set ratio. First, we used six different training set ratios, i.e., 10–60% with increments of 10% in the segmented layer, to implement stratified random sampling with the labeled objects, as described in Section 3.2.1. Objects with the remaining ratio of 40% were used as the validation set to calculate the OA. This process was repeated at scales from 10 to 200.

3.2.3. Features and feature selection

GEOBIA generates more features compared with pixel-based methods because it employs segmented objects. In the present study, we analyzed some features that are used frequently by eCognition software (eCognition Software® Definiens, 2011), including spectral measures, shape, and texture. These features were calculated using all three bands of the UAV data. The spectral

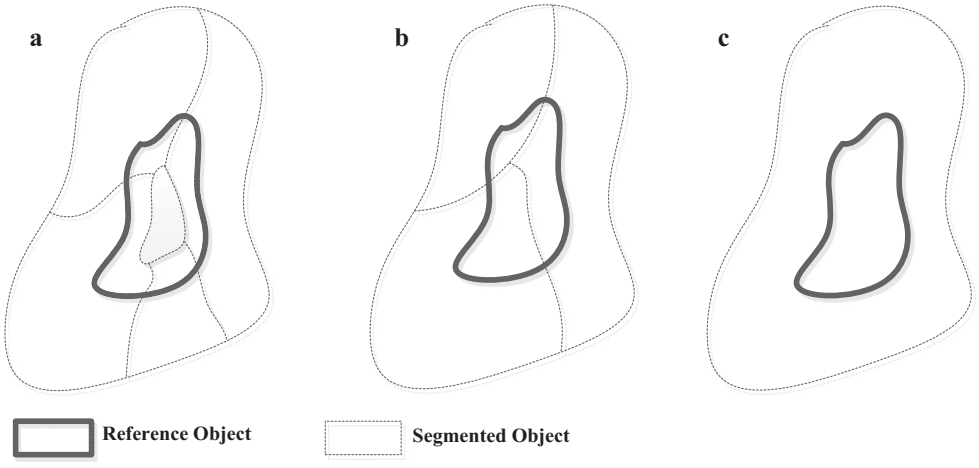

Fig. 3. Example showing the relationship between segmented objects (dotted line) and reference objects (heavy line). (a) The segmented objects contain the reference object at a fine scale, (b) they overlap at a medium scale, and (c) the reference object is contained within the segmented object at a coarse scale.

measures comprised mean blue, mean green, mean red, max difference, standard deviation blue, standard deviation green, standard deviation red, and brightness. The shape measures comprised area, compactness, density, roundness, main direction, rectangular fit, elliptic fit, asymmetry, border index, and shape index. The texture measures comprised GLCM (Gray-Level Co-occurrence Matrix) homogeneity, GLCM contrast, GLCM dissimilarity, GLCM entropy, GLCM std. dev., GLCM correlation, GLCM ang. 2nd moment, GLCM mean, GLDV (Gray-Level Difference Vector) ang. 2nd moment, GLDV entropy, GLDV mean, and GLDV contrast. The GLCM and GLDV were calculated based on the pixels in each object. They were used for each segmentation layer and in the feature selection algorithm.

In the present study, a best-first search strategy was used to obtain the optimal feature subset as well as feature importance evaluation using the gain ratio (Quinlan, 1996). First, we used the gain ratio as a measure of a single feature to rank all of the features in order to minimize the computational requirements. Second, the best-first search algorithm, which searches the space of the attribute subsets by greedy hill climbing augmented with a backtracking facility, was used to obtain the subset for assessment. Third, CFS was used to measure the quality of a subset of features, which were aggregated using the best-first search strategy based on the results of the feature importance evaluation.

The features ordered by the gain ratio at different scales are shown in Section 4.3. It is known that sorted features cannot represent the optimal subset for a classifier and that the feature subset with the top ranking cannot obtain the optimal classification results. The top ranking attributes were imported in advance by the best-first search algorithm for use as candidate features. In best-first search, two sets of nodes are maintained: open and closed. The open list is typically implemented using a priority queue, which is sorted by the gain ratio criteria, whereas the closed list is typically implemented as a hash table (Burns et al., 2010). A stopping criterion is used to prevent the best-first search from exploring the entire search space. If the evaluation criteria are not improved when the feature subset is added five consecutive times, the feature subset is no longer cycled to add more features.

CFS's feature subset evaluation function was used to select a subset of discriminative features based on the hypothesis that "a good feature subset contains features highly correlated with the class, yet uncorrelated with each other" (Hall and Holmes, 2003). The merit was calculated by Eq. (1):

$$\text{Merit} = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}} \quad (1)$$

where k is the number of features in the subset, "f" is the feature, "c" is the class, $\bar{r}_{cf}$ is the mean feature correlation with the class, and $\bar{r}_{ff}$ is the average feature inter-correlation. Both $\bar{r}_{cf}$ and $\bar{r}_{ff}$ are calculated using a measure based on conditional entropy (Press et al., 1988). The difference between a normal filter algorithm and CFS is that the normal filter provides scores for each feature independently, whereas CFS gives the heuristic merit of a feature subset and reports the best subset that it finds (Huang et al., 2008).

This procedure was performed automatically based on C# and Weka to output the features subset for the RF classifier at every scale. Furthermore, the optimal features for classification may be scale-dependent (Laliberte et al., 2012) for GEOBIA. Thus, CFS was applied at each scale before classification. Thus, the optimal feature subset for classification could be different at each scale.

3.2.4. RF classifier

Since the RF classifier was proposed (Breiman, 2001), it has been improved continuously in the field of remote sensing image information extraction, where it has been shown to be a robust classifier (Chan and Paelinckx, 2008; Rodriguez-Galiano et al., 2012; Puissant et al., 2014). RF employs a random method to establish a forest, which comprises many mutually independent decision trees. After obtaining the forest using the training set, we let each decision tree in the forest make a judgment about the unlabeled sample, before the unlabeled sample was predicted as the category that was voted for most frequently. Therefore, compared with other machine learning algorithms, RF can detect mislabeled examples of training information in GEOBIA that are associated with noise due to the essential sampling problem of GEOBIA.

The basic steps employed by RFs are as follows (Verikas et al., 2011).

(1) Each tree in a RF is grown using a bagging or bootstrap sample from the training set.
(2) When growing a tree, n variables (features) are selected randomly from the N available at each node (N is the number of the features imported).
(3) In general, $n \ll$ N. It is recommended to start with $n = \lfloor \log_2(N) + 1 \rfloor$ or $n = \sqrt{N}$, before decreasing and increasing n until the minimum error is obtained for the out of bag dataset. At each node, only the variable that provides the best split is used out of the n selected.

Thus, it is clear that the RF classifier only requires the definition of two parameters to generate a prediction model: the number of classification trees desired (k) and the number of prediction variables (n) used in each node to make the tree grow. When considering the parameter set used for remote sensing classification by the RF classifier, Rodriguez-Galiano et al. (2012) showed that it was preferable to use a large number of trees (k) and a small number of split variables (n) to reduce the generalization error and the correlation between trees. Based on their research results, our RF model used 479 trees and one single randomly split variable (feature) for UAV optical imagery classification. The overall classification procedure was performed automatically using C# and Matlab (RF classifier).

3.3. Accuracy assessment

The accuracy assessment method used for GEOBIA is considered to be an emerging research field because of its specific features compared with pixel-based classifiers (Blaschke, 2010; Whiteside et al., 2014; Radoux and Bogaert, 2014). Area-based methods have been used extensively for overall quality evaluations during building extraction (Freire et al., 2014). In general, area-based methods evaluate the classification accuracy in terms of the extent of features and their spatial distribution (Freire et al., 2014). GEOBIA has the potential to be affected by the scale problem, which means that the number of mixed objects will increase gradually and infinitely with the increase in scale, thereby leading to errors in object recognition. Obviously, these errors cannot be solved using a point-based accuracy assessment method, thus the area-based (polygon-based) accuracy assessment method is used to facilitate the production of a confusion matrix (Congalton and Green, 2009; Whiteside et al., 2014), which is used to calculate the User's Accuracy (UA), or commission error for each category. This suggests that the effect of the polygon's area on the classification accuracy should be considered during accuracy assessments for GEOBIA. Note that we also used the OA from the confusion matrix to evaluate the effect of the training dataset size (Section 3.2.2). To facilitate the analysis of scale effects on the classification accuracy, the accuracy assessment procedure was implemented automatically based on C# and ArcEngine 9.2.

4. Experiments and discussion

The aim of this study was to evaluate the effects of the training dataset size, feature selection, and scale on the classification accuracy based on a RF classifier. Fig. 4 shows the segmented layer obtained by MRS and the reference layer obtained by manual interpretation. Fig. 4a shows the segmentation results for area 1 at a scale of 100. Fig. 4b shows the manual interpretation layer for area 1, which comprised buildings, woodland, cropland, bare land, and road, where the mean area of each of these classes was 112.9, 1859.9, 557.8, 202.8, and 487.7 m^2, respectively. Fig. 4c shows the segmentation results for area 2 at a scale of 100. Fig. 4d shows the manual interpretation layer for area 2, which comprised buildings, woodland, cropland, bare land, road, and water, where the mean area of each of these classes was 213.3, 1105.4, 603.5, 160.9, 380.1, and 6316.4 m^2, respectively. The water category was not considered because there were too few samples after segmentation, especially at a coarse scale.

4.1. Evaluation of the training set size

To assess the effect of the training set size, we classified both images using the RF classifier to determine the OA. Fig. 4b and d show the manual interpretation layers and the segmented layers, where the manual interpretation results comprised cropland, woodland, buildings, water body, road, and bare land. First, we labeled all of the objects in the segmented layer based on the manual interpretation layer by intersecting the segmented layer with the reference layer. For example, we labeled a segmented object as cropland if it contained more than 60% cropland objects in the reference layer. The other segmented objects were labeled in a similar manner. Next, most of the segmented objects were given an attribute of category and stratified random sampling was implemented in the labeled segmented layer. To obtain the optimal value of the training set ratio, a number of experiments were performed using training set sizes with different ratios and different scales for segmentation. The training set ratio ranged between

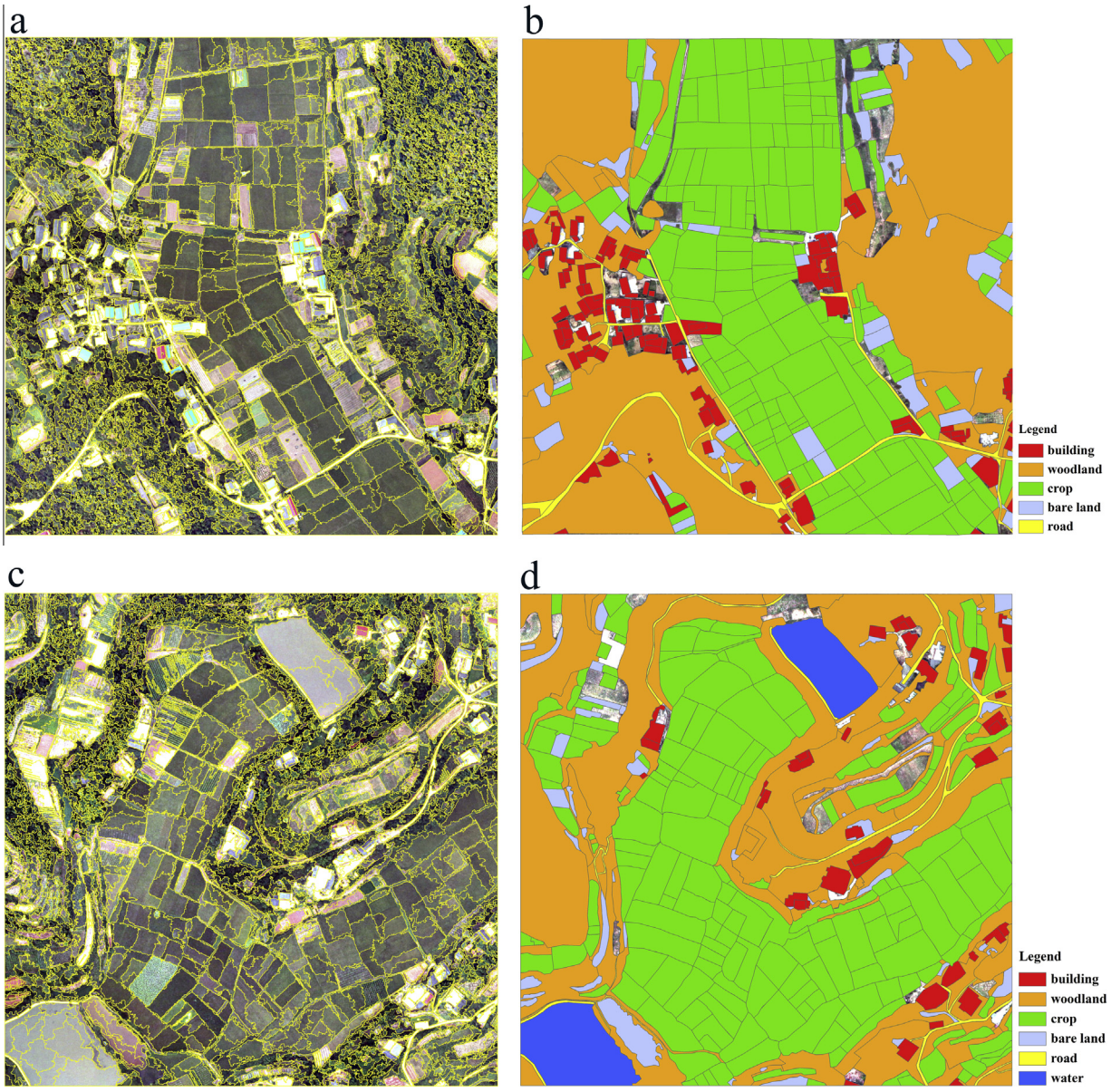

Fig. 4. The segmented layers and the corresponding reference layers. (a) The segmented layer for area 1 at a scale of 100; (b) the manual interpretation layer for area 1; (c) the segmented layer for area 2 at a scale of 100; and (d) the manual interpretation layer for area 2.

10% and 60%, and the scale of segmentation was fixed from 10 to 200, where we used intervals of 10% and 10, respectively. This yielded a total of 120 different classification results for each study area. The results were evaluated using the area-based OA (Congalton and Green, 2009; Radoux and Bogaert, 2014) with the 40% random test dataset. We only present the results for area 2 because the trends were the same (Fig. 5).

Fig. 5 shows the relationship between the segmentation scale and the training set ratio, which were used to train each RF model, and the classifier accuracy for the validation dataset. These results indicate that the level of accuracy increased with the size of the training set, whereas it declined as the segmentation scale increased, for GEOBIA. The same trends were found according to the pixel-based analysis, i.e., the OA was higher when the training set ratio was larger (Pal and Mather, 2003; Rodriguez-Galiano et al., 2012). In general, the training set with the optimal size should be used to analyze the effects of scale and features during UAV image classification. However, due to the limits of sampling, the maximum training samples we selected had a ratio of 60%. As the scale increased from 10 to 200, there was an unstable decline in the classification accuracy when the training set ratio was set at a lower level (i.e., 10%). There was also a stable declining trend when the training set ratio was set at a higher level (i.e., 50%). These results also indicate that there was high consistency in the OA at all 20 scales only when the training set ratio was high. According to these findings, we used a uniform training set ratio of 50% in the subsequent analysis.

The OA decreased as the scale increased, because under-segmentation increased as the scale increased, thereby yielding mixed objects. There is a natural error during classification due to under-segmentation (Kim et al., 2011; Witharana and Civco, 2014). Thus, an unstable decline may be more evident at a coarse scale, especially with a small training set ratio. For example, with a training set ratio of 10%, fluctuations began to occur when the scale parameter was >50; with a training set ratio of 20%, fluctuations began to occur when the scale parameter was >80; and with a training set ratio of 30%, fluctuations began to occur when the scale parameter was >120. With a training set ratio of 40%, the precision was relatively more unstable when the scale was >180, but when the training set ratio was 50%, the precision was more stable at all 20 scales. These results indicate that the accuracy assessment was only credible at a fine scale when the training set ratio was at a low level. For example, a scale parameter range of 10–60 could be used when the training set ratio is 10%. Consequently, the comparable range of the scale will also expand as the training set size increases.

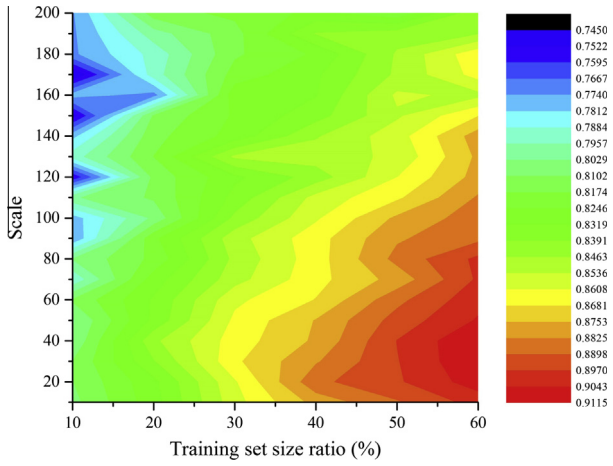

Fig. 5. Overall accuracy in terms of the relationship between the segmentation scale and the training set ratio.

By contrast, a training set ratio of 10% produced an OA of 0.8023 at a scale of 50, while a ratio of 30% generated an OA of 0.8625 at the same scale. However, the OA only increased to 0.8928 with a ratio of 50% for a scale of 50. Therefore, as the training set ratio increased from 10% to 30%, there was a rapid increase in the classification accuracy of 0.0602, whereas there was a slight increase of 0.0303 as the training set ratio increased from 30% to 50%. This indicates that the increase in the OA was slower with a large training set ratio. We found that the width of the contours in Fig. 5 increased markedly as the training set ratio increased, e.g., eight contours with ratios of 15–35%, but only four contours with ratios of 35–55%, where one contour represented a precision of approximately 0.0072 or 0.0073.

4.2. Sampling

To implement the stratified random sampling strategy at multiple scales, manual interpretations (Fig. 4b and d) of area 1 and area 2 were used as a reference layer to obtain a priori knowledge of the areas, which is a primary requirement of the stratified random sampling strategy (Congalton and Green, 2009). This part of the study did not focus on classification, but instead we aimed to explore the behavior of the training set size, features, scale when they varied. Thus, the land cover acquired by manual interpretation of the research area was used to overcome the primary problem of the stratified random sampling strategy. The reference was then overlapped with the segmented layer to provide the segmented object with a class label according to the rules described in Section 3.2.1. Subsequently, we used a training set ratio of 50% for each class based on the discussion in Section 4.1. It should be noted that Radoux and Bogaert (2014) also proposed rules for labeling different classes, e.g., 25% for urban and agriculture, and 75% for natural vegetation and water.

4.3. Gain ratio and CFS results

Figs. 6 and 7 show the relative contributions of the input features based on the gain ratio index for area 1 at scales of 60 and 110, respectively. These figures clearly show that the gain ratio values of the spectral features (F1–F8) and texture features (F19–F30) were significantly higher than that of the shape features (F9–F18), especially at a lower scale of 60. It should also be noted that the gap in the importance between shape features and spectral or texture features at a scale of 110 was less than that at a scale of 60. Thus, the effect of shape features may tend to become stronger as the scale increases. Moreover, the spectral and texture features estimated from scales of 60 and 110 had equal relative importance. Table 1 shows that a few texture features generated by CFS, including GLCM homogeneity, GLCM ang. 2nd moment, and GLCM mean, which are used frequently, contributed little of importance compared with the remaining features. This does not mean that the texture measures did not contribute to the classification of a specific category, but only that there was a strong correlation between the texture features and the spectral features because these measures were derived from the spectral information. In addition, we also present the results for the features subset filtered using CFS at scales of 60 and 110 (Figs. 6 and 7), which show that there was an optimal subset of 12 features at a scale of 60, whereas a subset of nine features comprised the most important variables for the classification at a scale of 110. Note that not all of the features in the top 12 or nine were imported into the optimal feature subset. For example, at a scale of 60, the spectral parameters with lower gain ratios, i.e., F4, F6, and F7, were considered the best subset, instead of F2 and F5. The same approach was applied to the textural and shape measures. The features with higher importance were ranked higher, but they were highly correlated with features that

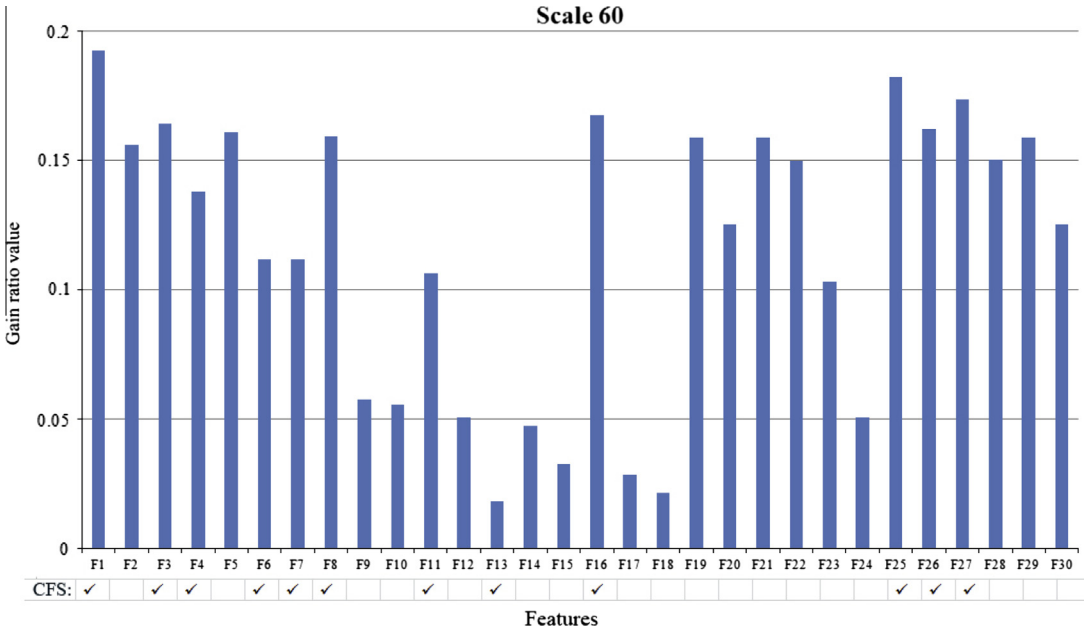

Fig. 6. Importance of features based on the gain ratio and the CFS results at a scale of 60 (area 1).

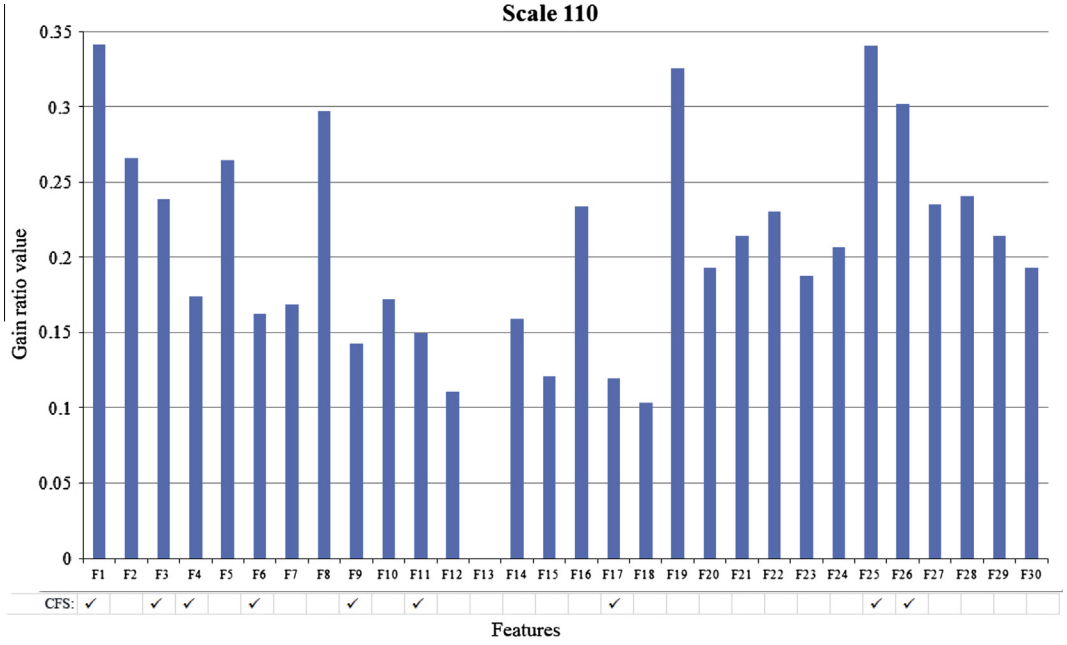

Fig. 7. Importance of features based on the gain ratio and the CFS results at a scale of 110 (area 1).

were more preferable, so they were not selected for the subsets. By contrast, the gain ratios of some features were low, but they were uncorrelated or weakly correlated with the other features. Therefore, they contributed more to the class separability. This also suggests that it is necessary to construct the optimal feature space in GEOBIA. Table 1 shows the optimal subsets obtained with different scales using the CFS method in area 1. In general, the results in this table suggest that F1, F3, F4, F6, F8, F11, F16, F19, F25, and F26 should be used as common features to avoid feature selection, and to calculate all of the features prior to the classification model.

4.4. Scale and classification accuracy

The scale-UA curves for each class were generated for areas 1 and 2. Fig. 8 (area 1) shows the UA behavior for three categories, i.e., cropland, woodland, and buildings, whereas bare land, roads, and water bodies are not listed because the sample sizes were too small to produce large fluctuations. In the same manner, Fig. 9 (area 2) shows the UA curves for cropland, woodland, and buildings, whereas bare land and roads are not shown. In area 1, it is clear that the UA was very stable for woodland at all scales (10–200), where a level of 95% was maintained without large fluctuations. However, the UA of cropland tended to decline between scales of 80 and 100, whereas it was almost invariable between scales of 10 and 80, but there was also an abrupt reduction when the scale exceeded 110. It should be noted that the UA of cropland decreased notably when the scale increased beyond 110. For the buildings category, we found that the UA tended to decrease when the scale was larger than 70 (Fig. 8). A similar behavior was observed in experimental area 2, with a declining UA trend for

Table 1
CFS results for area 1 at each scale.

ID	Features	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200
F1	**Mean blue**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
F2	Mean green																			✓	
F3	**Mean red**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
F4	**Max difference**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
F5	Standard deviation blue				✓						✓					✓	✓				
F6	Standard deviation green	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓				✓	✓	✓
F7	Standard deviation red	✓	✓	✓			✓			✓											
F8	**Brightness**	✓	✓	✓	✓		✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓
F9	**Area**	✓	✓	✓	✓	✓		✓	✓	✓		✓		✓		✓	✓	✓	✓	✓	✓
F10	Compactness									✓											
F11	**Density**			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
F12	Roundness																				
F13	Main direction	✓	✓	✓	✓	✓	✓	✓	✓												
F14	Rectangular fit																✓				
F15	Elliptic fit																				
F16	**Asymmetry**	✓	✓	✓	✓		✓	✓		✓	✓		✓	✓	✓	✓	✓			✓	
F17	Border index										✓	✓	✓	✓		✓		✓		✓	✓
F18	Shape index																✓		✓		
F19	**GLCM homogeneity**	✓	✓		✓			✓	✓	✓			✓	✓	✓		✓	✓			
F20	GLCM contrast	✓	✓										✓	✓		✓					
F21	GLCM dissimilarity																				
F22	GLCM entropy																				
F23	GLCM std. dev.																				
F24	GLCM correlation																				
F25	**GLCM ang. 2nd moment**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
F26	**GLCM mean**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				
F27	GLDV ang. 2nd Moment			✓		✓	✓														
F28	GLDV entropy																				
F29	GLDV mean										✓										✓
F30	GLDV contrast																				

Note: The best features for most scales are indicated in bold.

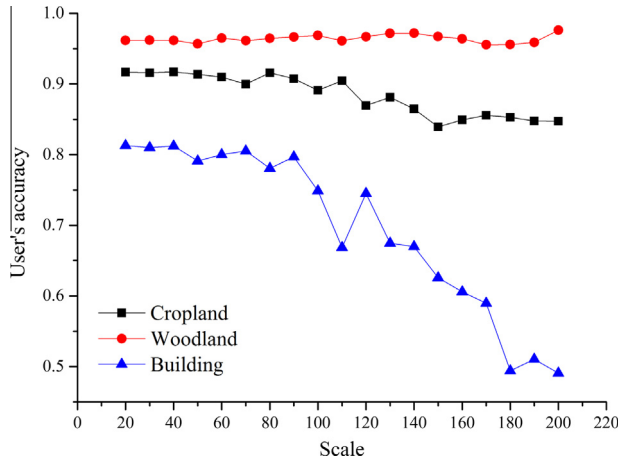

Fig. 8. Variations in the classification accuracy as the scale parameters increased using the RF classifier and area-based accuracy assessment method (area 1, CFS).

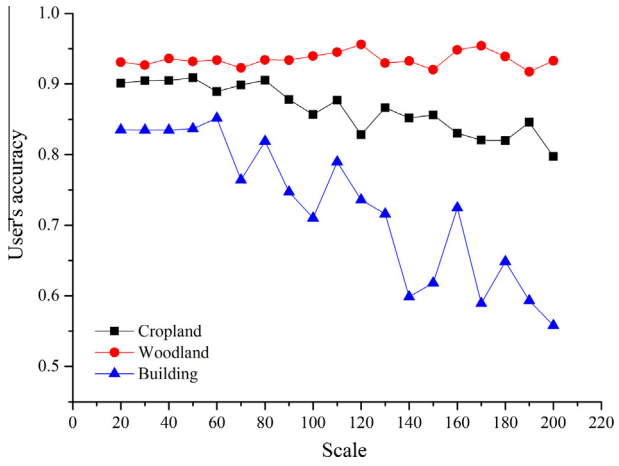

Fig. 9. Variations in the classification accuracy as the scale parameters increased using the RF classifier and area-based accuracy assessment method (area 2, CFS).

buildings above a scale of 60, while the decline for cropland occurred above a scale of 80 (Fig. 9). In addition, the threshold value for the decline was 170 for woodland based on the results with an extended scale range in area 2. A threshold value of 200 was obtained in the same manner for area 1.

The results described above show clearly that the UA of a specific map category, which did not decrease as the scale increased, was relatively stable at a fine scale. In other words, the UA did not increase with finer segmentation, but there were variable thresholds for different categories, after which there were significant declines as the scale increased. It should also be noted that our results do not agree with the findings of Laliberte and Rango (2009), who suggested that the UA increases as the scale increases. This may be attributable to the point-based accuracy evaluation method used in their early research. We tested the same accuracy assessment method in our experiments and we observed a similar trend to Laliberte and Rango (2009) (results not shown), i.e., an increase as the scale increased. However, this may be unreasonable because large polygons derived at a coarse scale are more likely to be under-segmented, thereby producing a large number of mixed objects that are less representative of their class. These polygons could then be misclassified because their mean spectral values do not reflect their actual content (Radoux and Bogaert, 2014). Thus, it should be noted that an accuracy assessment method that produces a sustained increase in accuracy is not encouraged for GEOBIA accuracy assessment.

For each class, the UA increased with the scale, which did not occur again with the area-based accuracy assessment method, but the UA declined abruptly at the start with different scales, which depended on the specific class. In addition to the UA behavior, the OA also decreased as the scale increased in both experiments, and the estimated OA results for area 2 are shown in Fig. 5 (the results for area 1 are not shown). Clearly, this is reasonable because more accurate information can be expressed by segmented objects at a finer scale. A segmented object (i.e., single cropland, building, and bare land areas that have been visually identified as single objects) has a small area at a very fine scale; thus, it cannot always represent the entire surface of an object completely, but it incorporates part of a true object with a lower risk of error compared with a coarse scale.

Figs. 8 and 9 show that the thresholds caused abrupt changes in the UA, which differed among the categories. In general, the UA-curve changed abruptly at this scale, after which the sum of the absolute values of UA difference between adjacent scales was greater than 5%. According to our analysis, the highest value was always obtained for woodland in both experiments, followed by the threshold for cropland and then buildings. However, the relationships between the thresholds of these categories were similar in both experiments. This is clearly reasonable because the largest area of an average single woodland is interpreted in the human layer. Although larger objects were generated as the scale increased, they still comprised the whole or part of a woodland object at a larger scale due to the essential large spatial regions, thereby avoiding mixed objects. For cropland, many mixed objects were produced at a large scale because the area of a single cropland was smaller than that of a woodland, which resulted in a lower UA level at a lower scale. Similarly, the average spatial areas of single buildings were smaller than those of both cropland and woodland; thus, the threshold for the reduction in the UA occurred at a relatively low scale, as shown in Figs. 8 and 9.

These results suggest a simple rule for the classification of remote sensing image using GEOBIA, i.e., we can select the smallest segmentation scale possible provided that we can ensure that the size of the training samples dataset is appropriate. In other words, over-segmentation (Kim et al., 2011) generally fails to reduce the classification accuracy by compensating for bias when using an adequate number of training samples, whereas under-segmentation (which leads to class mislabeling) has major effects on classification even if more training samples are used. In practice, it is necessary to determine an optimal segmentation scale to obtain better classification performance, where the threshold should be as high as possible to reduce the number of training samples required, as described in Section 4.1. At a fine scale, the number of segmented objects increases suddenly, which increases the computational load of the calculations, while the acquisition of true reference information (training set) for training the classifier is also a time-consuming task that incurs high economic costs.

Furthermore, it should be noted that the number of image objects obtained appears to depend on the SSP values identified (Drăguţ et al., 2014), which means that the average size of the actual objects determined by manual delineation may be proportional to the SSP values. Based on this hypothesis, we propose a semi-automatic method, which is combined with our previous analysis of the search for a threshold, thereby allowing the direct determination of the optimized scale parameters in various categories for a specific UAV image. To achieve this aim, the results of both experiments were analyzed comprehensively to elucidate the relationship between the optimal SSP value and the average size of the actual objects. First, the average sizes in different categories were calculated using both manual interpretation layers. Second, the optimal SSP values (previously called thresholds) were obtained that corresponded to the classes, as shown in Figs. 8 and 9. Fig. 10 shows a scatter plot of the optimal segmentation scale vs the mean area, which indicates that there was a strong positive linear correlation between the optimal segmentation scale and the mean area of actual objects. The correlation coefficient R^2 was 0.8628. Fig. 10 also shows that the size of the segmented objects was notably affected by the SSP, where the increasing size of the objects was associated with increasing SSP values (Blaschke, 2010; Witharana and Civco, 2014). Based on the linear fitting result for the scatter plot, we obtained an equation that represented the relationship between the optimal SSP and the average size of the real world objects based on the semantic interpretation: $y = 36.81 + 0.0939x$ (y is the optimal SSP and x is the mean area of real-world objects for each category in the reference layer). In practice, for specific UAV imagery, this equation suggests that the optimal SSP could be determined directly before classification using GEOBIA by referencing the general size of real-world objects, or by calculating the average size of the semantic interpretation of objects. For example, if the average size of the actual objects based on a human interpretation is 600 m^2 in a test area of cropland, the optimal SSP for the cropland is 90 according to the equation above.

In this study, only the single-scale classification results are shown in Fig. 11. The OA values obtained for areas 1 and 2 were 0.8943 at a scale of 110 and 0.8829 at a scale of 80, respectively (Tables 2 and 3). These results and our other considerations all suggest that a more efficient single-scale classification process can be obtained with a competitive OA by using the recommended scale rather than a coarse scale. According to the confusion matrix shown in Tables 2 and 3, the accuracies of the two areas are similar. In detail, the water in area 2 has the highest classification accuracy because of the stable and unique spectrum (Table 3), whereas there are many misclassifications of cropland and woodland in areas 1 and 2 because of their similar spectra. In addition, there were misclassifications of building and woodland/cropland because small buildings were always covered by trees, or they were close to cropland, which resulted in a high level of vegetation in some building pixels. Roads were differentiated with difficultly from the other categories, especially building and woodland. Bareland was easily confused with cropland and woodland, because fallow cropland and sparse woodland are similar to bareland.

4.5. A comprehensive analysis on effect of GEOBIA classification

Further analysis based on the previous results suggests that it is only necessary to use a lower training set ratio at a fine scale to obtain a higher level of classification accuracy, whereas a higher

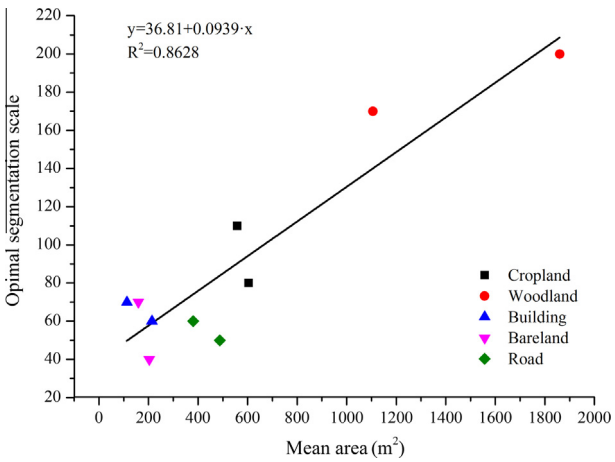

Fig. 10. Relationship between the optimal SSP (the scale was derived based on abrupt changes in the UA, as shown on the y-axis) and the average size or mean area of real-world objects based on the semantic interpretation (shown on the x-axis).

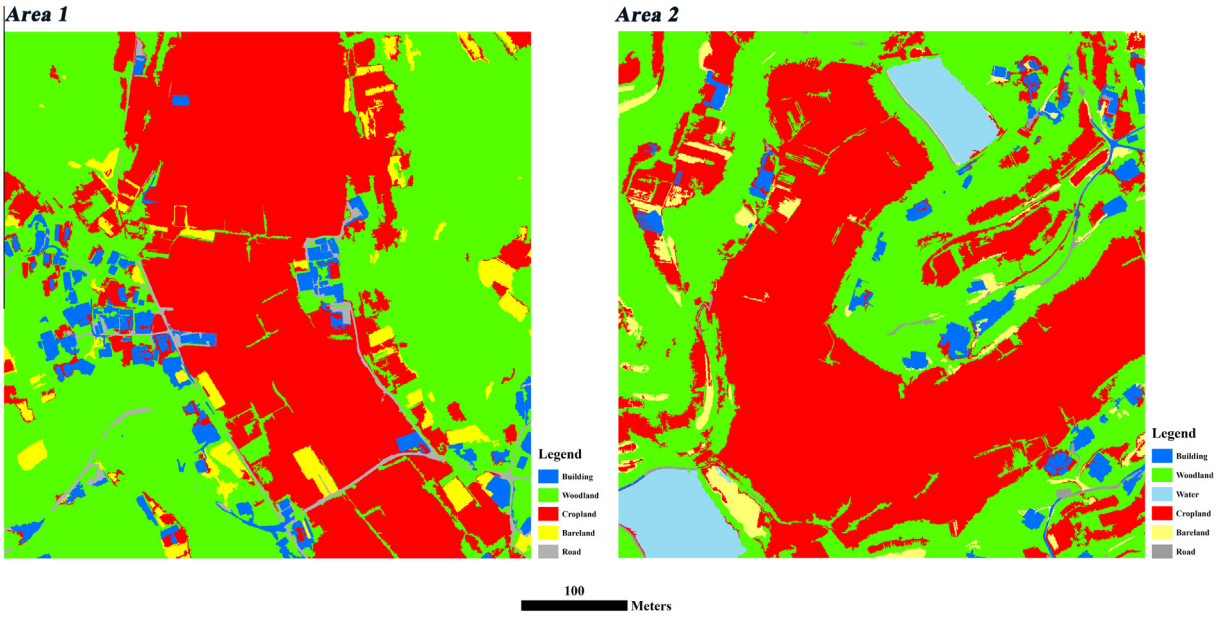

Fig. 11. Classification of experimental area 1 (left) and area 2 (right) with scale parameters of 110 and 80, respectively.

Table 2
Classification confusion matrix for area 1 at scale 110.

	Reference categories						
	Road	Bareland	Cropland	Woodland	Building	Sum (m^2)	Commission
Road	2527	162	251	954	978	4872	0.481
Bareland	252	8017	3062	1112	126	12,569	0.362
Cropland	265	2044	86,911	6973	226	96,419	0.099
Woodland	306	378	3071	103,607	473	107,835	0.039
Building	786	123	1625	1803	10,222	14,559	0.298
Sum (m^2)	4136	10,724	94,920	114,449	12,025		
Omission error	0.398	0.252	0.084	0.095	0.150		

Overall accuracy = 89.43%, kappa coefficient = 0.826.

Table 3
Classification confusion matrix for area 2 at scale 80.

	Reference categories							
	Road	Bareland	Cropland	Water	Woodland	Building	Sum (m^2)	Commission error
Road	1852	213	380	5	1045	303	3798	0.900
Bareland	83	5944	1904	0	948	158	9037	0.342
Cropland	15	808	100,909	0	11,875	452	114,059	0.115
Water	42	0	100	12,383	103	3	12,631	0.020
Woodland	148	804	6145	75	86,361	425	93,958	0.081
Building	106	141	957	0	1029	5656	7889	0.283
Sum (m^2)	2246	7910	110,395	12,463	101,361	6997		
Omission error	0.175	0.249	0.086	0.006	0.148	0.192		

Overall accuracy = 88.29%, kappa coefficient = 0.810.

training set ratio is required at a coarse scale. Obviously, this is common sense. Using the number of objects as a measure of the training set size for GEOBIA, the number of objects increases sharply at a fine scale, which leads directly to the requirement for a fairly large training set to represent the classes adequately, despite the lower sampling ratio. By contrast, the number of objects decreases sharply at a coarse scale for the same area and the training set is always too small to represent the classes adequately for classification, despite the higher sampling ratio. In general, the classification accuracy can be enhanced by increasing the training set size for pixel-based methods or GEOBIA. However, GEOBIA is more complex because of its unfixed segmentation scale; thus, the effect of the training set size in GEOBIA needs to be analyzed urgently, especially the effect of different combinations on the scale.

In terms of feature analysis, the spectral features have obvious advantages, although the textural and shape features should not be ignored, but they also require more attention. There have been many previous studies of textural measures (Laliberte and Rango, 2009; Kim et al., 2011; Duro et al., 2012; Laliberte et al., 2012) but only three features were selected frequently for every scale in our test, i.e., the GLCM homogeneity, GLCM ang. 2nd moment, and GLCM mean. This indicates that most of the textures may be notably correlated with the spectral measures. In summary, more

measures provide more information for classification in GEOBIA, but the spectra are still most important. According to Table 1, we also recommend that the shape features of density, asymmetry, and area should receive more attention, and the border index needs to be considered at a coarse scale. During object-based analysis, it is time consuming to compute texture measures (Laliberte and Rango, 2009); thus, we should try to control the number of texture features when producing applications.

As discussed previously, we also found that the SSP values were proportional to the average area of real-world objects of interest and various categories corresponded to the different SSP values. Clearly, the general sizes of real-world objects based on prior knowledge (i.e., the size of cropland and buildings) can be used directly, instead of the average size of interpretation categories, to determine the optimal SSP value for each category. Multiscale classification is a current developmental trend in the use of GEOBIA for processing VHR remote sensing images (Kim et al., 2011; Kurtz et al., 2014). There is no doubt that the optimal SSP for each category is useful for multiscale classification, and most of these methods were designed to identify one optimal segmentation rather an optimal multi-scale set of segmentations (Johnson and Xie, 2013). However, the optimal SSP values obtained using our method may cause a problem in selecting a uniform scale when a single-scale classification is employed. Nevertheless, we cautiously recommend that the optimal SSP for a category, which considers the total large area, is a practical scale parameter. For example, the optimal SSP values were determined from cropland as a scale of 110 in area 1 and a scale of 80 in area 2, and the results were reliable (Tables 2 and 3), but without the need for complicated multiscale classification.

5. Conclusions

Using area-based accuracy assessment methods, the main aim of the present study was to assess the optimal SSP for GEOBIA, particularly for information extraction from VHR optical images obtained by UAV, and to develop a semi-automatic method for determining the optimized scale parameter. Our specific objectives were to study the behavior of the classification accuracy with changes in the size of the training dataset and the scale parameter variables. We also investigated feature measures at multiple scales in GEOBIA, where CFS was shown to be a useful algorithm for selecting suitable feature measures to facilitate computation. The results of this study provide new insights into the use of GEOBIA for UAV optical imagery information extraction.

The accuracy of the classification produced using GEOBIA was notably affected by changes in the training set ratio and the scale parameter. When the SSP was fixed, the accuracy increased with the training set ratio in the same manner as pixel-based classification (Pal and Mather, 2003; Rodriguez-Galiano et al., 2012). This was a predictable outcome, but there was a decline in the classification performance of the multiscale model as the SSP value increased, which was attributable to the production of hybrid objects with the gradual increase in the size of the segmented objects. The results also indicated that the increase in the OA declined with a high training set ratio. Furthermore, we also found that a training set with a smaller ratio was not appropriate for classification at a coarse scale; thus, a large training set ratio is necessary to obtain reliable classification accuracy at a coarse scale, whereas a lower ratio is suitable for obtaining the same level of performance at a fine scale. For example, the same OA of 0.8681 was achieved at scales of 40 and 110 with the recommended training set ratios of 30% and 50%, respectively. It is worth to mention that the large ratios (i.e., 50%) should be not recommended because it is not appropriate in any practical use. The small ratios may be recommended more in fine scales for the practical use.

In MRS-based GEOBIA, the results generally suggest that a smaller scale yields a better OA given an adequate training set. Thus, the classification accuracy is assumed to be better when the scale is smaller. However, if the objects are segmented at a fine scale (over-segmentation) and there are similarities among different real objects, they are also difficult to identify even if sufficient training samples are used. Thus, we focused on determining the appropriate scale for generating segmented objects that represented real objects. Objects have different sizes and various categories correspond to different optimal SSP values; thus, the determination of the optimal SSP is a pre-condition for hierarchical information extraction (Kurtz et al., 2014). Our results showed that the average size of the actual objects based on manual delineation was proportional to the SSP values. Thus, we obtained an equation for determining the optimal SSP for specific UAV imagery based on the actual objects acquired by manual delineation, where this equation was always valid for specific imagery data. Furthermore, we developed a semi-automatic method for the direct determination of the optimized scale parameter for various categories. This equation is expected to be applicable to UAV image data analysis.

The present study differs from single information extraction using UAV images (Laliberte and Rango, 2009; Lucieer et al., 2014). UAV are viable platforms for land cover monitoring, especially in complex landscapes, where CFS and the RF classifier in GEOBIA are suitable for classifying UAV imagery. Given the recent increased use of high-resolution digital aerial cameras in piloted aircraft, our results are also relevant to this area. In summary, the results of the present study contribute to our understanding of the applications of UAV remote sensing imagery and GEOBIA, where both fields will remain innovative and vital for surface information extraction for many years to come.

We also recognize that this study has potential limitations. First, we used the RF parameters acquired by Rodriguez-Galiano et al. (2012), which were tested based on pixels from Landsat Thematic Mapper-5 data. A further test with GEOBIA should be performed to enhance the RF model. Second, area-based methods are different from point-based methods when assessing the accuracy in GEOBIA and further assessments are required. Meanwhile, it should be emphasized that this study focused on UAV images, which necessarily have a relatively small size. The findings of this research therefore apply specifically to small images classified with GEOBIA. Nevertheless, the findings may have implications for larger images, a subject which we will address in future research.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant Nos. 41371017 and 41001238), the National Key Technology R&D Program of China (Grant Nos. 2012BAH28B02). Sincere thanks are given for the comments and contributions of anonymous reviewers and members of the editorial team.

References

Addink, E.A., Van Coillie, F., De Jong, S.M., 2012. Introduction to the GEOBIA 2010 special issue: from pixels to geographic objects in remote sensing image analysis. Int. J. Appl. Earth Obs. Geoinf. 15, 1–6.

Arvor, D., Durieux, L., Andrés, S., Laporte, M.A., 2013. Advances in Geographic Object-Based Image Analysis with ontologies: a review of main contributions and limitations from a remote sensing perspective. ISPRS J. Photogram. Remote Sens. 82, 125–137.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J. Photogram. Remote Sens. 58 (3), 239–258.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogram. Remote Sens. 65 (1), 2–16.

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., et al., 2014. Geographic Object-Based Image Analysis-Towards a new paradigm. ISPRS J. Photogram. Remote Sens. 87, 180–191.
Breiman, L., 2001. Random forests. In: Schapire, R.E. (Ed.), Machine Learning. Springer, Netherlands, 5–32.
Burns, E., Lemons, S., Ruml, W., Zhou, R., 2010. Best-First heuristic search for multicore machines. J. Artific. Intell. Res. 39, 689–743.
Chan, J.C., Paelinckx, D.E., 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sens. Environ. 112 (6), 2999–3011.
Cheng, L., Gong, J., Yang, X., Fan, C., Han, P., 2008. Robust affine invariant feature extraction for image matching. IEEE Geosci. Remote Sens. Lett. 5 (2), 246–251.
Cheng, L., Li, M., Liu, Y., Cai, W., Chen, Y., Yang, Kang, 2012. Remote sensing image matching by integrating affine invariant feature extraction and RANSAC. Comput. Electr. Eng. 38 (4), 1023–1032.
Cheng, L., Tong, L., Liu, Y., Li, M., Wang, J., 2014. Automatic registration of coastal remotely sensed imagery by affine invariant feature matching with shoreline constraint. Mar. Geodesy 37 (1), 32–46.
Colomina, I., Molina, P., 2014. Unmanned aerial systems for photogrammetry and remote sensing: a review. ISPRS J. Photogram. Remote Sens. 92, 79–97.
Congalton, R.G., Green, K., 2009. Assessing the accuracy of remotely sensed data: principles and practices. CRC/Taylor and Francis Group, LLC, Boca Raton, London, New York.
Costa, H., Carrao, H., Bacso, F., Caetano, M., 2014. Combining per-pixel and object-based classifications for mapping land cover over large areas. Int. J. Remote Sens. 35 (2), 738–753.
Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. ISPRS J. Photogram. Remote Sens. 84, 85–99.
Drăguţ, L., Tiede, D., Levick, S., 2010. ESP: a tool to estimate scale parameters for multiresolution image segmentation of remotely sensed data. Int. J. Geograph. Inform. Sci. 24 (6), 859–871.
Drăguţ, L., Csillik, O., Eisank, C., Tiede, D., 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. ISPRS J. Photogram. Remote Sens. 88, 119–127.
Dunford, R., Michel, K., Gagnage, M., Piegay, H., Tremelo, M.L., 2009. Potential and constraints of unmanned aerial vehicle technology for the characterization of mediterranean riparian forest. Int. J. Remote Sens. 30 (19), 4915–4935.
Duro, D.C., Franklin, S.E., Dube, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sens. Environ. 118 (15), 259–272.
Foody, G.M., Mathur, A., Sanchez-Hernandez, C., Boyd, D.S., 2006. Training set size requirements for the classification of a specific class. Remote Sens. Environ. 104 (1), 1–14.
Freire, S., Santos, T., Navarro, A., Soares, F., Silva, J.D., Afonso, N., et al., 2014. Introducing mapping standards in the quality assessment of buildings extracted from very high resolution satellite imagery. ISPRS J. Photogram. Remote Sens. 90, 1–9.
Gomez-Candon, D., De Castro, A.I., Lopez-Granados, F., 2014. Assessing the accuracy of mosaics from unmanned aerial vehicle (UAV) imagery for precision agriculture purposes in wheat. Precision Agric. 15 (1), 44–56.
Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.
Hall, M.A., Holmes, B., 2003. Benchmarking attribute selection techniques for discrete class data mining. IEEE Trans. Knowledge Data Eng. 15 (6), 1–16.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. ACM SIGKDD Explor. Newslett. 11 (1), 10–18.
Hay, G.J., Castilla, G., 2008. Geographic Object-Based Image Analysis (GEOBIA): a new name for a new discipline. In: Blaschke, T., Lang, S., Hay, G.J. (Eds.), Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications. Springer, Berlin, Heidelberg, pp. 75–89.
Huang, C.J., Yang, D.X., Chuang, Y.T., 2008. Application of wrapper approach and composite classifier to the stock trend prediction. Expert Syst. Appl. 34 (4), 2870–2878.
Hussain, M., Chen, D.M., Cheng, A., Wei, H., Stanley, D., 2013. Change detection from remotely sensed images: from pixel-based to object-based approaches. ISPRS J. Photogram. Remote Sens. 80, 91–106.
Jaakkola, A., Hyyppa, J., Kukko, A., Yu, X., Kaartinen, H., Lehtomäki, M., et al., 2010. A low-cost multi-sensoral mobile mapping system and its feasibility for tree measurements. ISPRS J. Photogram. Remote Sens. 65 (6), 514–522.
Johnson, B., Xie, Z., 2011. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. ISPRS J. Photogram. Remote Sens. 66 (4), 473–483.
Johnson, B., Xie, Z., 2013. Classifying a high resolution image of an urban area using super-object information. ISPRS J. Photogram. Remote Sens. 83, 40–49.
Kim, M., Warner, T.A., Madden, M., Atkinson, D.S., 2011. Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. Int. J. Remote Sens. 32 (10), 2825–2850.
Kim, J., Lee, S., Ahn, H., Seo, D., Park, S., Choi, C., 2013. Feasibility of employing a smartphone as the payload in a photogrammetric UAV system. ISPRS J. Photogram. Remote Sens. 79, 1–18.
Kurtz, C., Stumpf, A.E., Malet, J., Gancarski, P., Puissant, A., Passat, N., 2014. Hierarchical extraction of landslides from multiresolution remotely sensed optical images. ISPRS J. Photogram. Remote Sens. 87, 122–136.
Laliberte, A.S., Rango, A., 2009. Texture and scale in Object-based analysis of subdecimeter resolution Unmanned Aerial Vehicle (UAV) imagery. IEEE Trans. Geosci. Remote Sens. 47 (3), 761–770.
Laliberte, A.S., Browning, D.M., Rango, A., 2012. A comparison of three feature selection methods for object-based classification of sub-decimeter resolution UltraCam-L imagery. Int. J. Appl. Earth Obs. Geoinf. 15, 70–78.
Liu, Y.X., Li, M.C., Mao, L., Xu, F., Huang, S., 2006. Review of remotely sensed imagery classification patterns based on object-oriented image analysis. Chin. Geograph. Sci. 16 (3), 282–288.
Lucieer, A., Turner, D., King, D.H., Robinson, S.A., 2014. Using an Unmanned Aerial Vehicle (UAV) to capture micro-topography of Antarctic moss beds. Int. J. Appl. Earth Obs. Geoinf. 27, 53–62.
Ma, L., Li, M.C., Tong, L.H., Wang, Y.F., Cheng, L., 2013a. Using unmanned aerial vehicle for remote sensing application. In: 2013 21st International Conference on Geoinformatics, Kaifeng, China, pp. 1–5.
Ma, L., Wang, Y.F., Li, M.C., Tong, L.H., Cheng, L., 2013b. Using High-Resolution Imagery Acquired with an Autonomous Unmanned Aerial Vehicle for Urban Construction and Planning. In: International Conference on Remote Sensing, Environment and Transportation Engineering (RSETE 2013), Najing, China.
Ma, L., Cheng, L., Han, W.Q., Zhong, L.S., Li, M.C., 2014. Cultivated land information extraction from high-resolution unmanned aerial vehicle imagery data. J. Appl. Remote Sens. 8 (1), 83673.
Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q.H., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sens. Environ. 115 (5), 1145–1161.
Novack, T., Esch, T., Kux, H., Stilla, U., 2011. Machine learning comparison between WorldView-2 and QuickBird-2-Simulated imagery regarding Object-Based urban land cover classification. Remote Sens. 3 (10), 2263–2282.
Pal, M., Mather, P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sens. Environ. 86 (4), 554–565.
Pal, M., Foody, G.M., 2010. Feature selection for classification of Hyperspectral data by SVM. IEEE Trans. Geosci. Remote Sens. 48 (5), 2297–2307.
Powers, R.P., Hay, G.J., Chen, G., 2012. How wetland type and area differ through scale: a GEOBIA case study in Alberta's Boreal Plains. Remote Sens. Environ. 117 (15), 135–145.
Press, W.H., Flannery, B.P., Teukolsky, S., Vetterling, W.T., 1988. Numerical recipes. Cambridge University Press, Cambridge, UK.
Puissant, A., Rougier, S., Stumpf, A.E., 2014. Object-oriented mapping of urban trees using Random Forest classifiers. Int. J. Appl. Earth Obs. Geoinf. 26, 235–245.
Quinlan, J.R., 1996. Improved use of continuous attributes in C4.5. J. Artific. Intell. Res. 4, 77–90.
Radoux, J., Bogaert, P., 2014. Accounting for the area of polygon sampling units for the prediction of primary accuracy assessment indices. Remote Sens. Environ. 142, 9–19.
Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogram. Remote Sens. 67, 93–104.
Stehman, S.V., Wickham, J.D., 2011. Pixels, blocks of pixels, and polygons: choosing a spatial unit for thematic accuracy assessment. Remote Sens. Environ. 115 (12), 3044–3055.
Stefanski, J., Mack, B., Waske, B.O.R., 2013. Optimization of object-based image analysis with random forests for land cover mapping. IEEE J. Selected Topics Appl. Earth Observ. Remote Sens. 6 (6), 2492–2504.
Sugiura, R., Noguchi, N., Ishii, K., 2005. Remote-sensing technology for vegetation monitoring using an unmanned helicopter. Biosyst. Eng. 90 (4), 369–379.
Trimble Documentation eCognition® Developer 8.7 Reference Book, München, Germany, 22 September 2011.
Van Niel, T.G., McVicar, T.R., Datt, B., 2005. On the relationship between training sample size and data dimensionality: Monte Carlo analysis of broadband multi-temporal classification. Remote Sens. Environ. 98 (4), 468–480.
Verikas, A., Gelzinis, A., Bacauskiene, M., 2011. Mining data with random forests: a survey and results of new tests. Pattern Recogn. 44 (2), 330–349.
Whiteside, T.G., Maier, S.W., Boggs, G.S., 2014. Area-based and location-based validation of classified image objects. Int. J. Appl. Earth Obs. Geoinf. 28, 117–130.
Witharana, C., Civco, D.L., 2014. Optimizing multi-resolution segmentation scale using empirical methods: exploring the sensitivity of the supervised discrepancy measure Euclidean distance 2 (ED2). ISPRS J. Photogram. Remote Sens. 87, 108–121.
Zhang, C.H., Kovacs, J.M., 2012. The application of small unmanned aerial systems for precision agriculture: a review. Precision Agric. 13 (6), 693–712.
Zhang, X.L., Xiao, P.F., Song, X.Q., She, J.F., 2013. Boundary-constrained multi-scale segmentation method for remote sensing images. ISPRS J. Photogram. Remote Sens. 78, 15–25.
Zhen, Z., Quackenbush, L.J., Stehman, S.V., Zhang, L., 2013. Impact of training and validation sample selection on classification accuracy and accuracy assessment when using reference polygons in object-based classification. Int. J. Remote Sens. 34 (19), 6914–6930.
Zhou, G., Ambrosia, V., Gasiewski, A.J., Bland, G., 2009. Foreword to the special issue on Unmanned Airborne Vehicle (UAV) sensing systems for earth observations. IEEE Trans. Geosci. Remote Sens. 47 (3), 687–689.