ISPRS Journal of Photogrammetry and Remote Sensing xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Classifying a high resolution image of an urban area using super-object information

Brian Johnson *, Zhixiao Xie

Department of Geosciences, Florida Atlantic University, 777 Glades Rd., Boca Raton, FL 33431, United States

ARTICLE INFO

Article history:
Received 12 December 2011
Received in revised form 10 May 2013
Accepted 16 May 2013
Available online xxxx

Keywords:
Segmentation
Classification
Urban
High resolution
Land cover
Scale
Contextual

ABSTRACT

In this study, a multi-scale approach was used for classifying land cover in a high resolution image of an urban area. Pixels and image segments were assigned the spectral, texture, size, and shape information of their super-objects (i.e. the segments that they are located within) from coarser segmentations of the same scene, and this set of super-object information was used as additional input data for image classification. The accuracies of classifications that included super-object variables were compared with the classification accuracies of image segmentations that did not include super-object information. The highest overall accuracy and kappa coefficient achieved without super-object information was 78.11% and 0.727%, respectively. When single pixels or fine-scale image segments were assigned the statistics of their super-objects prior to classification, overall accuracy increased to 84.42% and the kappa coefficient increased to 0.804.

1. Introduction

Urban land cover information extracted from high resolution aerial or satellite imagery can be used for a variety of purposes, including urban tree canopy mapping (Walton et al., 2008), green space mapping (Lang et al., 2008), impervious surface mapping (Zhou and Wang, 2008), and updating building footprint GIS data (Jin and Davis, 2005). Land cover data can also be helpful for mapping urban land use (Herold et al., 2003). However, extracting land cover information from high resolution data can be difficult when traditional pixel-based image classification methods are used due to the high degree of spectral variability within land cover classes (caused by shadows, sun angle, gaps in tree canopy, etc.) that causes low classification accuracy (Yu et al., 2006). This high degree of within-class spectral variability is due to the fact that a single pixel typically represents only a small part of a classification target (e.g. tree canopy, building rooftop, or road) in a high-resolution image. The mismatch between pixels and real-world objects of interest is related to the modifiable areal unit problem (MAUP; Openshaw, 1984), and occurs because pixels in remote sensing images are arbitrary in size and thus typically do not correspond well with real-world objects. In several previous studies, a geospatial object-based image analysis (GEOBIA or simply OBIA; Blaschke, 2010) approach, in which an image is segmented into relatively homogeneous regions (i.e. "segments" or "image objects") prior to classification, has outperformed the pixel-based approach (Thomas et al., 2003; Blaschke et al., 2004; Yu et al., 2006; Myint et al., 2011). In the OBIA method, the attributes of these segments are used for classification instead of attributes of single pixels. Use of segments rather than single pixels as the base units for analysis can reduce within-class spectral variability because representative values of segments (e.g. mean values) are used instead of individual pixel values. It also allows for spatial and contextual information such as size, shape, texture, and topological relationships (e.g. containment and adjacency) to be incorporated for classification (Blaschke et al., 2004; Benz et al., 2004). Finally, image segments are less sensitive to MAUP than pixels because they better match the objects of interest in the image (Hay and Castilla, 2006). However, the object-based approach is not without problems.

One issue with the object-based approach is that classification accuracy is affected by image segmentation quality (Liu and Xia, 2010). Some image segmentation algorithms, such as the "Multi-resolution Segmentation" region-merging algorithm described by Benz et al. (2004), require users to set one or more parameters that influence the average size of segments produced by a segmentation (i.e. the segmentation scale), and choosing appropriate parameters can be difficult due to the fact that a single set of parameters can produce very different segmentation results depending on the properties of the imagery (e.g. bit depth, number of bands, spatial resolution, image heterogeneity) (Dragut et al., 2010). Choosing

* Corresponding author. Tel.: +1 561 297 3250; fax: +1 561 297 2745.
E-mail address: bjohns53@fau.edu (B. Johnson).

0924-2716/$ - see front matter © 2013 Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS).
http://dx.doi.org/10.1016/j.isprsjprs.2013.05.008

segmentation parameter(s) that produce segments smaller than the actual ground features in an image results in over-segmentation, which is undesirable because non-spectral information (e.g. size and shape) calculated for segments will not be useful for classification. Using parameter(s) that produce segments larger than the actual features in an image results in under-segmentation, which is undesirable because segments will contain pixels from more than one type of land cover. Over-segmentation and under-segmentation have both been shown to lower classification accuracy, but the effect of under-segmentation is generally considered to be worse (Kim et al., 2009; Liu and Xia, 2010). To deal with this scale issue, it is common to employ a multi-scale classification approach that involves building a hierarchy of multiple segmentations and classifying different types of land cover at each segmentation scale based on expert knowledge (e.g. Burnett and Blaschke, 2003; Dorren et al., 2003; Zhou and Troy, 2009; Myint et al., 2011; Kim et al., 2011). However, manually choosing appropriate segmentation scale(s) and decision rules to classify each type of land cover requires a detailed investigation of segments at each scale, and the procedure can be time intensive and subjective. For more automated classification tasks (e.g. supervised classification using training samples), to minimize over- and under-segmentation some studies have compared multiple segmentations of a scene prior to classification to identify the best one (Kim et al., 2008; Trias-Sanz et al., 2008), while other have identified the optimal segmentation after classification by comparing the classification accuracies of each segmented image (Dorren et al., 2003; Kim et al., 2010; Liu and Xia, 2010). The main downside to these single-scale supervised classification approaches is that different types of land cover may be classified better at different scales, so using only one segmentation scale may not produce the best results. A multi-scale supervised classification approach that does not require users to investigate segments and develop classification rules at each scale would be faster and less subjective.

One possible method for classifying different types of land cover at different scales comes from the fact that some segmentation algorithms, such as the Multiresolution Segmentation algorithm mentioned previously, produce a hierarchy in which segments generated at a fine scale are nested inside of segments generated at coarser scales, as shown in Fig. 1. The larger segments are referred to as super-objects of the smaller segments, and the smaller segments are referred to as sub-objects of the larger segments (Definiens, 2006). Spectral and non-spectral information (e.g. size and shape) of these super-objects may be useful for image classification purposes. For example, spectral information from the smallest segments may be useful for targeting individual trees, while size and shape information from larger super-objects may be useful for separating buildings from concrete and other surfaces spectrally similar to building rooftops. Since the main problem with the single-scale classification approach is that not all types of land cover are segmented best at one scale, the theoretical advantages of this multi-scale approach are that: (i) including multiple segmentation scales for classification makes it more likely that at least one of them corresponds well to each type of land cover, and (ii) use of multi-scale variables provides information on how each type of land cover behaves across many scales rather than at only one scale. Another advantage of this approach is that it is less reliant on expert knowledge and less subjective than traditional multi-scale classification methods that require segments to be investigated at every segmentation scale in order to determine the best scale(s) for classifying each type of land cover.

Fig. 1. Segment generated at a fine scale (border shown as a grey line) located within a segment generated at a coarser scale (border shown as a black line).

A previous study by Bruzzone and Carlin (2006) found that the use of contextual information of pixels (i.e. the spectral and non-spectral values of the segments that contain a pixel) increased the accuracy of a pixel-based land cover classification. However, in their study, classification accuracy was not compared with that achieved when a single-scale OBIA supervised classification approach (segmentation followed by classification) was used. Since, as previously stated, the object-based approach has been shown to outperform the pixel-based approach for classification of high resolution imagery, it is possible that using a single-scale OBIA approach will still work better than a pixel-based method that incorporates super-object information. For this reason, it is necessary to compare the classification accuracy achieved when pixels are classified using super-object information with the accuracy achieved using a single-scale OBIA approach in order to see which yields better results. Furthermore, it is necessary to compare a pixel-based approach that incorporates super-object information with an object-based approach that incorporates super-object information (i.e. with segments as the base units for analysis) to see which, if either, is preferable for classifying high resolution imagery.

In this study, we will: (1) compare the classification accuracies achieved when (a) super-object information is included with (b) the classification accuracies achieved using traditional single-scale supervised classification methods, and (2) test the use of both single pixels and image segments as the basic units for the super-object classification. Although this is not the first study to use super-object information for classification (Bruzzone and Carlin (2006) also used it), our systematic investigation of these single-scale and multi-scale supervised classification approaches should be useful for future OBIA research.

We consider a large number of spectral and non-spectral variables for classification (up to 153 for the pixel-based classification that includes super-object information, 20 for the single-scale segmentations), and some statistics may be correlated, so a classification algorithm that can handle high dimensional datasets containing some redundant variables is needed. The random forest algorithm proposed by Breiman (2001) was chosen for this study because it has been shown to perform well for classifying hyperspectral images (Ham et al., 2005; Lawrence et al., 2006), which also contain many input variables and redundant variables. The random forest classifier is an ensemble classifier that uses a random subset of the input variables and a bootstrapped sample of the training data to perform a decision tree classification (Breiman, 2001). Typically, a large number of trees are generated, and

Please cite this article in press as: Johnson, B., Xie, Z. Classifying a high resolution image of an urban area using super-object information. ISPRS J. Photogram. Remote Sensing (2013), http://dx.doi.org/10.1016/j.isprsjprs.2013.05.008

unweighted voting is used to determine final class assignments for each pixel or image segment. Some advantages of the random forest classifier are its speed, relative insensitivity to user-defined parameters, insensitivity to noise and overtraining, and ability to achieve results comparable to classifiers that are more computationally intensive (e.g. boosting) or require more parameter calibration (e.g. support vector machines) (Breiman, 2001; Pal, 2005; Gislason et al., 2006).

2. Study area and data

For this study, 30 cm resolution color infrared (CIR) digital aerial orthoimagery of the city of Deerfield Beach, Florida, USA (26°18′40″N, 80°5′48″W), was obtained from the Broward County Property Appraiser (flight date: December 31, 2008). The imagery contains 8-bit data for the near infrared (NIR), red, and green spectral bands. For our study area, we chose a 4630 pixel × 4967 pixel (approximately 1400 m × 1500 m) subset that contains many of the types of land cover typically found in an urban area, including buildings and other sealed surfaces (concrete and asphalt), trees, shrubs, grass, swimming pools, and bare soil. There are also vehicles (cars, trucks, and boats) and shadows present in the image. Since there was only one small lake in our study area, we chose to mask it out rather than include it for image classification purposes. The color infrared image of the study area is shown in Fig. 2.

3. Methods

3.1. Image segmentation

Image segmentation was performed in Definiens Professional 5 (currently with Trimble) using the Multiresolution Segmentation algorithm, which starts with one-pixel image segments, and merges neighboring segments together until a heterogeneity threshold is reached (Benz et al., 2004). The heterogeneity threshold is determined by a user-defined "scale parameter", as well as color/shape and smoothness/compactness weights. In general, increasing the value of the scale parameter causes the average size of segments to increase.

For this study, a series of image segmentations was performed using seven different scale parameters (20–140 at an interval of 20) so that the image could be analyzed at several scales. Scale parameters smaller than 20 produced segments that were, in general, over-segmented relative to all of the land cover classes of interest in this study, and parameters larger than 140 produced segments that were under-segmented relative to all land cover

Fig. 2. Color infrared image of the study area.

Please cite this article in press as: Johnson, B., Xie, Z. Classifying a high resolution image of an urban area using super-object information. ISPRS J. Photogram. Remote Sensing (2013), http://dx.doi.org/10.1016/j.isprsjprs.2013.05.008

types of interest. Segments produced using three different scale parameters are shown in Fig. 3 to allow for a visual comparison. The scale parameter interval of 20 was chosen in order to keep the number of super-objects to a reasonable level, while still adequately capturing the features in the image at different scales. All three spectral bands were assigned equal weights for segmentation because all of them may contain useful information. Color/shape weights were set to 0.9/0.1 because we wanted spectral information to have the most important role in the segmentation, and smoothness/compactness weights were set to 0.5/0.5 because we did not want to favor either compact or non-compact segments.

For each segment, spectral information (mean values and variance for each band, mean normalized difference vegetation index (NDVI)), texture information (gray-level co-occurrence matrix (GLCM) contrast, correlation, and entropy for the NIR band (all directions)), and size/shape statistics (area, roundness, shape index, border index, length/width, rectangular fit, density, border length, and asymmetry) were calculated. We used the NIR band for GLCM texture calculations because it is the most useful spectral band for classifying vegetation land cover types, and it is also useful for non-vegetation land cover. In choosing the number of variables to use for image classification, we chose to err on the side of using too many variables rather than too few, since the random forest algorithm is capable of handling high dimensional datasets and is not very sensitive to redundant variables. However, equations for each of the variables were investigated prior to choosing them to ensure that they were not too similar. For more information about the formulas used for texture and size/shape calculations, readers are encouraged to refer to the Definiens Professional 5 Reference book (Definiens, 2006).

3.2. Assigning super-object statistics to segments

Once the statistics for each segment were calculated, a series of spatial joins was performed in ArcGIS so that the smallest segments (i.e. segments generated using a scale parameter of 20) could be assigned the spectral, texture, and size/shape statistics of their super-objects as well. The end result was that segments generated using a scale parameter of 20 also contained the statistics of their super-objects generated using scale parameters from 40 to 140. For the sake of comparison, we also used this process to assign super-object information to single pixels (in addition to the spectral values and NDVI of the pixels).

3.3. Image classification

In this section, land cover classification is performed for: (a) image segments containing super-object statistics and (b) single pixels containing super-object statistics. Classification is also performed for the seven segmentations that do not include super-object statistics in order to assess the impact that using super-object information has on classification. Finally, we perform a per-pixel classification without super-object information to allow for a comparison with all of the other classifications. An overview of the image classification workflow is shown in Fig. 4.

To classify the segmentations that did not include super-object statistics, training data were collected for each type of land cover from segments generated using a scale parameter of 20, referred to from now on as the "scale 20 segmentation". A total of 168 segments were used as training data for classification. Super-objects of these training segments were used as the training data for the other segmentations (i.e. super-objects from the scale 40 segmentation were used as training data to classify the scale 40 segmentation, and so on). As an example, training segments for two different segmentation scales are shown in Fig. 5. For the pixel-based classification, one pixel within each of the training segments was chosen as a training pixel.

For the classifications that included super-object information, spatial joins were performed, as described in Section 3.2, so that the fine-scale training segments (or training pixels) could be assigned the statistics of their super-objects. Because, as discussed in the Introduction section, using segments larger than the actual features of interest (i.e. under-segmented image objects) results in lower classification accuracy, we tested classification accuracy as super-object information from each of the coarser segmentations was added to see if the accuracy decreased when information from the coarsest segmentations was included for classification. For example, we performed classification when super-object information from the scale 40 segmentation was assigned to the scale 20 segmentation, then again when the scale 60 information was also added, and so on. Since over-segmentation also affects classification accuracy, we tested different base units for classification as well (e.g. single pixels, scale 20 segments, and scale 40 segments). When the scale 20 segments were used as the base units, the values of single pixels were not included for classification, and when the scale 40 segments were used as the base units, values of single pixels and scale 20 segments were not included for classification.

Reference data, used for accuracy assessment, were collected using a stratified systematic unaligned sampling scheme (Jensen,

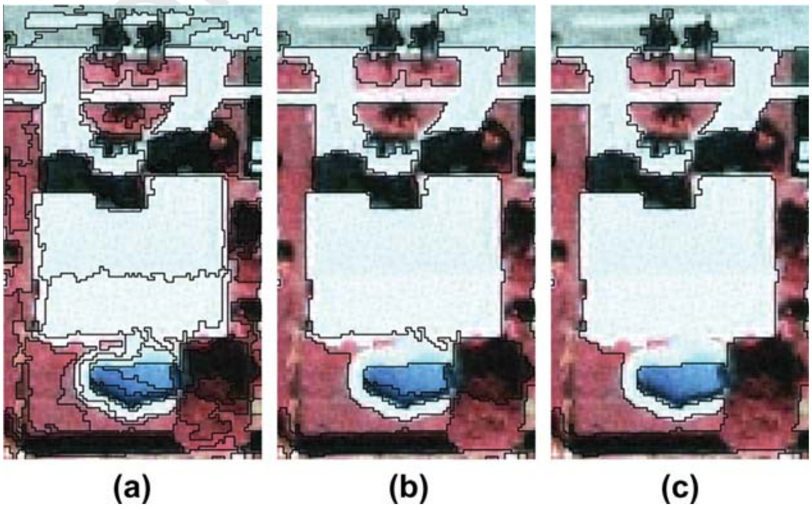

Fig. 3. Scale 20 (a), 80 (b), and 140 (c) segments overlaid on the color infrared imagery for a subset of the image.

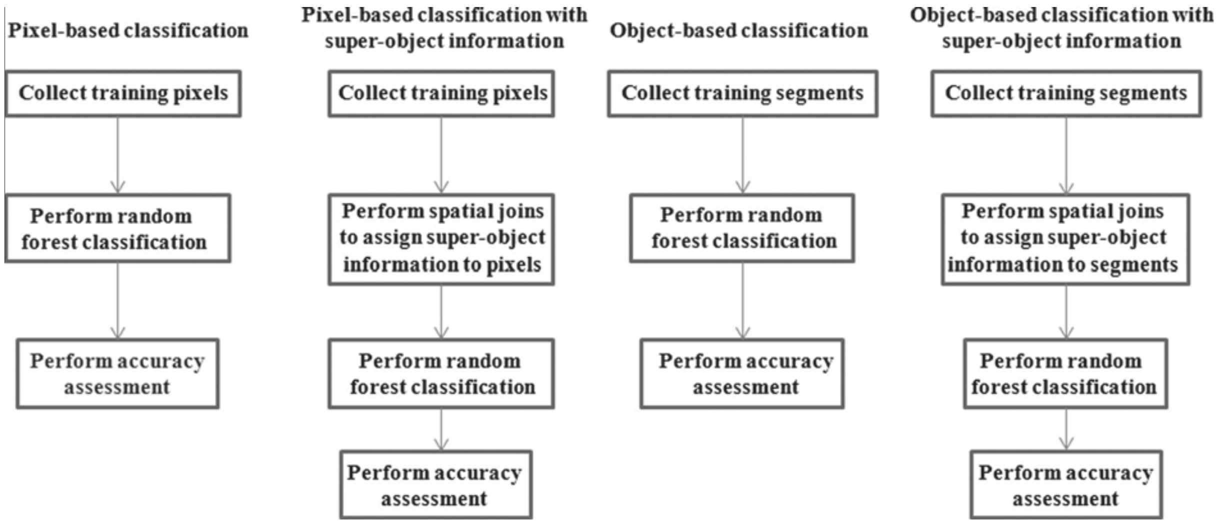

Fig. 4. Flowchart of the classification methods used in this study.

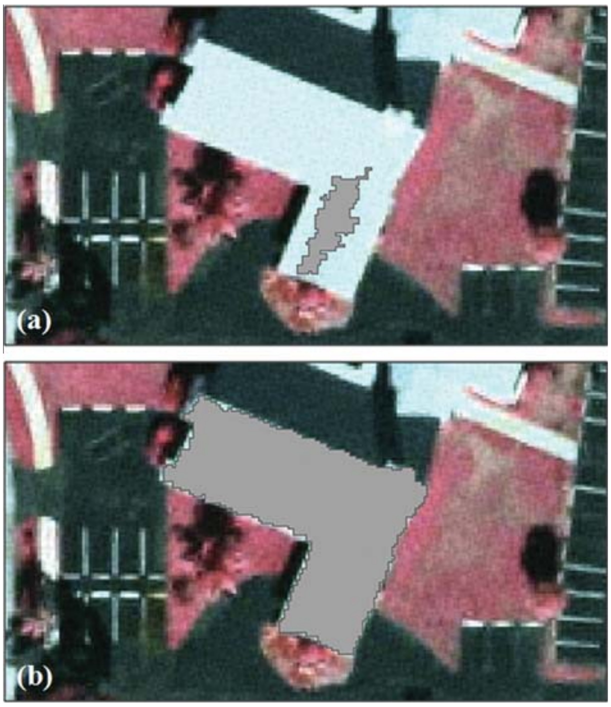

Fig. 5. "Building" class training segment for the scale 20 segmentation (a) and the scale 100 segmentation (b). For the scale 20 classifications that included super-object information, scale 20 training segments were also assigned the information of their super-objects.

2005). A grid consisting of 100 m × 100 m cells was overlaid on the image, and within each cell 3 random segments were chosen from the scale 20 segmentation. This sampling method ensured that the test data were randomly located, yet distributed across the entire image. In total, the reference data consisted of 507 segments, with segments being assigned to a land cover class based on visual interpretation of the high-resolution imagery. Super-objects of the randomly-selected reference segments were used as reference data for the remaining segmentations (i.e. super-objects from the scale 40 segmentation were used as reference data for the scale 40 segmentation, and so on). Reference data for the segmentations that included super-object statistics were created by once again performing spatial joins. A random point was created inside each of the reference segments to select reference pixels for the pixel-based classification that did not include super-object information.

Random Forest classification was performed using Weka 3.6.4, an open-source data mining program (Hall et al., 2009). There were two user-defined parameters required to perform classification: the number of decision trees to create and the number of randomly-selected variables considered for splitting each node in a tree. Previous research has shown that the number of trees and the number of randomly-selected variables selected have a relatively small impact on classification accuracy (Breiman, 2001; Pal, 2005). Breiman (2001) reported good results for datasets of different sizes when the number of variables was set to $\log_2 M + 1$, where M is the number of variables, and Lawrence et al. (2006) found that using 500 trees or more produced unbiased estimates of error. Based on the previous research, we set the number of trees as 500, and the number of variables used for splitting each node as $\log_2 M + 1$. These settings minimized the computation time for each classification because optimum parameters did not need to be identified, which was important for this study because a large number of classifications were performed.

4. Results and discussion

4.1. Classification system

Image segments (or pixels for the pixel-based classifications) were classified into the following land cover classes: grass, trees/shrubs, buildings, concrete, asphalt, vehicles, bare soil, and pools. In some previous studies (Bruzzone and Carlin, 2006; Walker and Blaschke, 2008), buildings were split up into more than one class when training data were collected (white roof, red tile roof, etc.) so that spectral information would be more useful for classifying buildings. However, in our study area, rooftops were so diverse in terms of color and building materials that this was not practical. Instead, we used one building class that included rooftops of different colors, and relied more on non-spectral information for classifying buildings correctly. After classification, segments classified as concrete, asphalt, and vehicles were aggregated into a single land cover class called "other impervious" because concrete and asphalt are both impervious surfaces, and vehicles are likely to be located on top of an impervious surface.

4.2. Overall classification accuracy

Overall accuracies for many of the classifications are shown in Fig. 6. The highest overall accuracy (84.42%) and kappa coefficient (0.804) were achieved when either pixels or scale 20 segments were also assigned super-object information from the scale 40, 60, and 80 segmentations for classification. For a pixel-based classification that did not include super-object information, overall accuracy (73.18%) and the kappa coefficient (0.666) were much lower. For the single-scale segmentations, overall accuracy was very similar when scale parameters between 20 and 60 were used, with the scale 40 segmentation achieving the highest overall accuracy (78.11%) and kappa coefficient (0.727). As the scale parameter was increased past 60, overall accuracy of the single-scale segmentations decreased as segments started to consist of pixels from more than one land cover class. This decrease in overall accuracy as the image became more and more under-segmented is consistent with results found in past studies (Kim et al., 2010; Liu and Xia, 2010). Table 1 shows the error matrices for: the pixel-based classification, the optimal single-scale segmentation/classification, the pixel-based classification using super-object information, and the scale 20 segmentation/classification using super-object information. To test whether or not the increase in classification accuracy achieved with super-object information was statistically significant, pairwise *z*-tests (Congalton, 2009) were calculated to compare the error matrices produced with and without super-object variables. The null hypothesis of each pairwise *z*-test is that the two error matrices being compared have no significant difference. Based on the *z* scores, reported in Table 2, the error matrices were statistically different at a 95% confidence level ($\alpha = 0.05$), confirming that super-object variables contributed significantly to improvement in classification accuracy. Fig. 7 shows a subset of the aerial imagery, the classified scale 20 segments with super-object information from scale 40, 60, and 80 segmentations, and the optimal single-scale segmentation/classification (i.e. the scale 40 segmentation) to allow for a visual comparison.

In all of the cases that we tested, the use of super-object information improved overall accuracy. However, accuracy decreased slightly from its highest levels when super-object information from the coarsest segmentations (scale 100, 120, 140 segmentations) was included for classification due to the fact that ground features surrounded by spectrally similar land cover (e.g. trees surrounded by grass, buildings surrounded by concrete/asphalt) were under-segmented in the coarse segmentations. Based on these results, we recommend using super-object information for classification purposes, but care should be taken to avoid using super-object information from segmentations that are highly under-segmented. We also observed a decrease in overall accuracy when larger segments were used as the base units for classification. When the base units were changed from scale 20 segments to scale 40 segments, the highest overall accuracy that was achieved decreased to 82.64% from 84.42%, and the kappa coefficient decreased to 0.782 from 0.804. This decrease in accuracy occurred because small features, such as single trees, became under-segmented. We also tested the use of scale 60 segments and scale 80 segments as the base units for classification. The highest overall accuracy for the scale 60 segments (80.08%) and highest kappa coefficient (0.745) was achieved when they were assigned super-object information from scale 80, 100, and 120 segments, and the highest overall accuracy (77.31%) and kappa coefficient (0.714) for the scale 80 segments were achieved when they were assigned super-object information

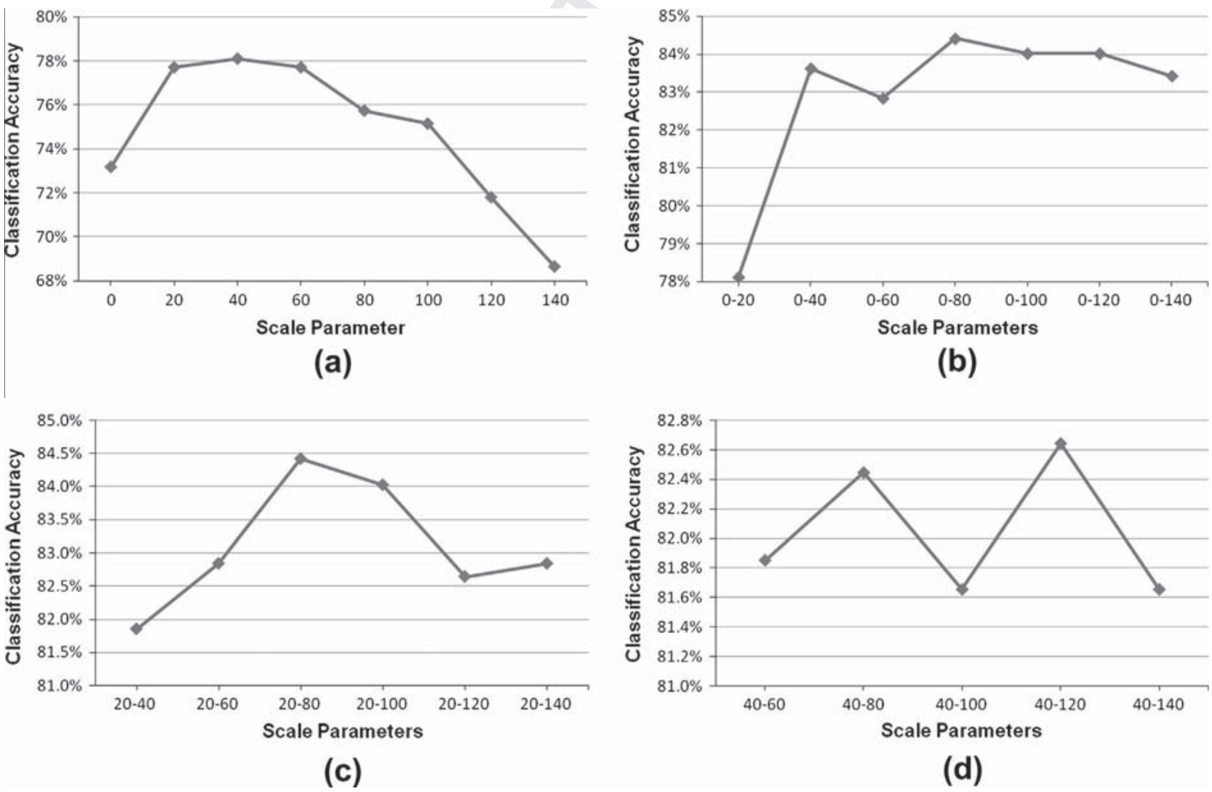

Fig. 6. Overall classification accuracies for the pixel-based and single-scale classifications (a), and the multi-scale classifications with: single pixels (b), scale 20 segments (c), and scale 40 segments (d) as the base units for classification. Scale parameter of 0 indicates a pixel-based classification. For the multi-scale classifications, "Scale Parameters" indicate which segmentations were used for classification (e.g. "Scale Parameters" of 0–80 indicate that super-object variables from the scale 20, 40, 60, and 80 segmentations was assigned to single pixels for classification).

Please cite this article in press as: Johnson, B., Xie, Z. Classifying a high resolution image of an urban area using super-object information. ISPRS J. Photogram. Remote Sensing (2013), http://dx.doi.org/10.1016/j.isprsjprs.2013.05.008

Table 1
Error matrices for the optimal single-scale classification (a), the optimal scale 20 classification with super-object variables (b), the pixel-based classification (c), and the optimal pixel-based classification with super-object variables (d). *Note*: G, grass; T, tree; B, building; I, other impervious; SH, shadow; SO, soil; P, pool; PA, producer's accuracy; UA, user's accuracy.

Reference data	G	T	B	I	SH	SO	P	Total	PA (%)
(a) Classification data									
G	76	5	1	0	0	1	0	83	91.57
T	26	58	0	0	5	0	0	89	65.17
B	0	0	71	22	0	2	2	97	73.20
I	1	0	25	124	3	3	3	159	77.99
SH	0	0	0	3	42	0	0	45	93.33
SO	4	0	2	1	0	13	0	20	65.00
P	1	0	1	0	0	0	12	14	85.71
Total	108	63	100	150	50	19	17	507	
UA (%)	70.37	92.06	71.00	82.67	84.00	68.42	70.59		
Overall accuracy									78.11%
Kappa									0.727
(c) Classification data									
G	72	6	0	1	1	3	0	83	86.75
T	35	49	0	0	5	0	0	89	55.06
B	0	0	48	38	0	11	0	97	49.48
I	0	1	20	127	6	5	0	159	79.87
SH	0	0	0	1	44	0	0	45	97.78
SO	0	0	0	0	0	20	0	20	100.00
P	1	0	2	0	0	0	11	14	78.57
Total	108	56	70	167	56	39	11	507	
UA (%)	66.67	87.50	68.57	76.05	78.57	51.28	100.00		
Overall accuracy									73.18%
Kappa									0.666

Reference data	G	T	B	I	SH	SO	P	Total	PA (%)
(b) Classification data									
G	75	6	1	1	0	0	0	83	90.36
T	12	73	0	0	4	0	0	89	82.02
B	0	0	72	23	0	2	0	97	74.23
I	1	1	10	140	4	1	2	159	88.05
SH	0	0	0	1	44	0	0	45	97.78
SO	4	0	1	3	0	12	0	20	60.00
P	1	0	1	0	0	0	12	14	85.71
Total	93	80	85	168	52	15	14	507	
UA (%)	80.65	91.25	84.71	83.33	84.62	80.00	85.71		
Overall accuracy									84.42%
Kappa									0.804
(d) Classification data									
G	76	6	1	0	0	0	0	83	91.57
T	14	71	0	0	4	0	0	89	79.78
B	0	0	70	23	0	4	0	97	72.16
I	0	1	12	141	3	2	0	159	88.68
SH	0	0	0	0	45	0	0	45	100.00
SO	1	0	2	4	0	13	0	20	65.00
P	1	0	1	0	0	0	12	14	85.71
Total	92	78	86	168	52	19	12	507	
UA (%)	82.61	91.03	81.40	83.93	86.54	68.42	100.00		
Overall accuracy									84.42%
Kappa									0.804

Table 2
Pairwise comparison of error matrices for classifications performed with and without super-object variables. Scale Parameter "40" is the most accurate single-scale classification, "20–80" is the most accurate scale 20 segmentation with super-object variables, "0" is the pixel-based classification, and "0–80" is the most accurate pixel-based classification with super-object variables.

Scale parameter(s)	*Z* score	α Value	Significant at α = 0.05?
40 vs. 20–80	2.52	0.01	Yes
0 vs. 0–80	4.37	0.00	Yes
0–80 vs 20–80	0.004	1.00	No

from the scale 100 and 120 segments. To assess the statistical significance between the classifications performed with different base classification units, we performed the pairwise *z*-tests shown in Table 3. In this table, it is clear that there was a decrease in accuracy as larger and larger segments are used as the base units, and this decrease became significant at a 95% confidence level (α = .05) when the scale 60 or coarser segments are used as the base units. Due to this trend, we did not perform classification using scale 100 or larger segments as the base units. Based on our results, we recommend that single pixels or very fine-scale segments (i.e. seg-

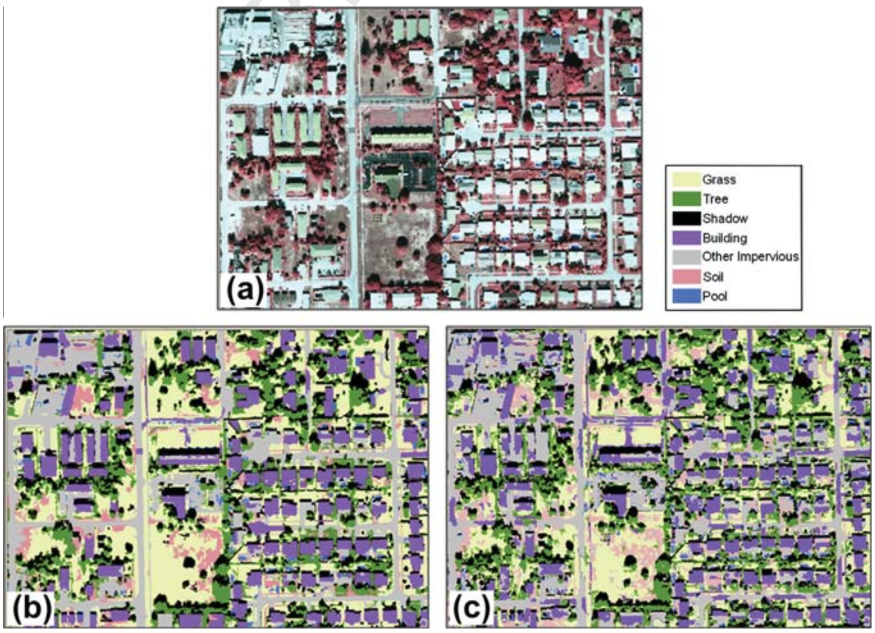

Fig. 7. Subset of the study area image (a), and land cover maps produced from the most accurate multi-scale (b) and single-scale (c) classifications. In general, there is a better correspondence between the imagery and the multi-scale classification.

Please cite this article in press as: Johnson, B., Xie, Z. Classifying a high resolution image of an urban area using super-object information. ISPRS J. Photogram. Remote Sensing (2013), http://dx.doi.org/10.1016/j.isprsjprs.2013.05.008

Table 3
Pairwise comparison of error matrices for classifications that included super-object variables, performed using different scales of segments as the base units. "20–80" is the most accurate scale 20 segmentation with super-object variables, "40–120" is the most accurate scale 40 segmentation with super-object variables, and "60–120" is the most accurate scale 60 segmentation with super-object variables, and "80–120" is the most accurate scale 80 segmentation with super-object variables.

Scale Parameters	*Z* score	α Value	Significant at α = 0.05?
40–120 vs. 20–80	0.75	0.45	No
60–120 vs. 20–80	1.95	0.05	Yes
80–120 vs. 20–80	2.91	0.00	Yes

ments no larger than any of the land cover types of interest) be used as the base units for a classification that includes super-object information, as the effect of over-segmentation had less of a negative impact on classification accuracy than under-segmentation (super-object information reduces errors caused by over-segmentation but not under-segmentation).

Another factor related to segmentation scale that may have had an impact on our classification accuracies was that, like many other OBIA studies (e.g. Liu et al., 2010; Johnson and Xie, 2011; Kim et al., 2011), we used a regular interval for the selecting the scale parameters. This means that our choices were somewhat arbitrary and possibly not optimal for the land cover of interest. Due to the huge number of possible segmentation parameter settings, parameter optimization is very challenging. Some OBIA studies have used quantitative approaches to identify optimal segmentations, either by comparing image segments with manually-digitized reference polygons (Marpu et al., 2010) or using empirical "goodness measures" that mimic human perception of a good segmentation (Kim et al., 2008; Johnson and Xie, 2011). Most of these methods were designed to identify one optimal segmentation rather an optimal multi-scale set of segmentations. However, a multi-scale segmentation optimization tool called the Estimation of Scale Parameter (ESP) tool, recently developed by Dragut et al. (2010), allows for more than one optimal segmentation to be identified. Although we did not use the ESP tool in this study (it was not compatible with our version of Definiens Professional), a possible future research topic would be to evaluate the impact that ESP's parameter optimization has on the accuracy of a multi-scale supervised classification (compared to classification using arbitrarily-chosen of segmentation parameters).

4.3. Accuracy by land cover class

For the pixel-based and scale 20 classifications that included super-object variables from scale 40, 60, and 80 segmentations, most classes achieved producer's and user's accuracies of 80% or better. As shown in Table 1, the producer's and user's accuracies of most land cover classes were also improved when super-object variables were included for classification. The error matrices of segments with super-object information and pixels with super-object information are very similar ($z = 0.004$), but for land cover classes for which shape information is useful (i.e. "Building" and "Tree"), slightly higher producer's and user's accuracies were achieved when segments were used as the base units for classification. For these classes, the spectral information of pixels was highly variable (due to different roof top colors and shadows within roof tops/tree canopies), leading to lower classification accuracies. For classes without regular shapes (e.g. "Grass, "Soil", and "Other impervious"), accuracy was slightly higher when pixels were used as the base units, so we can infer that the pixel information was useful to some degree for these classes.

Fig. 8 shows the land cover map of the entire study area produced when scale 20 segments were classified using super-object variables from scale 40, 60, and 80 segmentations. Visual comparison of the study area imagery in Fig. 2 and the classified map in Fig. 8 shows a relatively good correspondence. However, due to the spectral similarity between buildings and other land covers such as "Soil", "Buildings", and "Other impervious", there were often some misclassifications for these classes. For example, "Soil" segments on some baseball fields had shapes similar to buildings in the study area, so they were misclassified as "Building". It was also clear that, within the "Building" class, buildings surrounded by vegetation (most single-family houses) were classified correctly more often than buildings surrounded by spectrally-similar land cover (e.g. concrete or asphalt). This occurred because, after segmentation, some segments that contained pixels of a building also contained pixels of the spectrally-similar land cover surrounding the building, causing the segments' shapes to be inaccurate.

Use of additional datasets for image segmentation and classification, such as Light Detection and Ranging (LIDAR) height data, would likely lead to fewer classification errors for buildings and trees surrounded by spectrally-similar land covers. However, while multispectral aerial imagery of the study area is typically acquired annually by the Broward County Property Appraiser's Office, LIDAR acquisition is rarer (LIDAR imagery is only available for 2004 and 2007). To assess the level of overall accuracy that could be achieved on an annual basis, we used only multispectral imagery in this study.

4.4. Relationship between classified segments and real-world objects

As previously discussed, we found that the use of spectral and non-spectral variables from super-objects of pixels or fine-scale image segments led to higher overall accuracy than when variables from a single scale were used for classification. However, the limitation we encountered with using pixels/small segments as the base units for classification was that they were smaller than most of the real-world objects of interest. For example, a tree or building often consisted of several segments or pixels, rather than just one. For this reason, we emphasize that the pixels and image segments classified using the methods described in this study contained accurate categorical information (i.e. class assignment), but they did not have a good one-to-one relationship with real-world objects of interest. We were only interested in the accuracy of the categorical information in this study because for our application

Fig. 8. Classified map of the study area, produced by classifying scale 20 segments with super-object variables from scale 40, 60, and 80 segments.

Please cite this article in press as: Johnson, B., Xie, Z. Classifying a high resolution image of an urban area using super-object information. ISPRS J. Photogram. Remote Sensing (2013), http://dx.doi.org/10.1016/j.isprsjprs.2013.05.008

(creating an urban land cover map) it was not important whether a land cover object consisted of one or multiple classified segments, as long as the segments were assigned to the correct class. For that reason, like similar OBIA studies (e.g. Bruzzone and Carlin, 2006; Liu et al., 2010; Kim et al., 2011), we used the classical categorical accuracy assessment measures originally developed for pixel-based classifications (e.g. producer's accuracy, user's accuracy, etc.). In contrast, some OBIA studies have also considered the spatial accuracy of image segments (i.e. the accuracy of segment boundaries) in the accuracy assessment procedure (e.g. Tiede et al., 2010). Including measures that also quantify the spatial accuracy of segment boundaries can be very useful for some applications, particularly when it is important to get accurate counts and/or area estimates of the objects of interest (since under-segmentation or over-segmentation would distort the counts and area estimates). For example, accurate segment boundaries would be important for a study that aimed to identify the number of trees or buildings in an image, or estimate each tree's canopy size/each building's footprint. When pixels or small segments are the base units for classification, this type of analysis is not possible without further processing. Merging adjacent pixels or segments of the same land cover class together (e.g. merging neighboring "Building" segments into a single building) may provide more accurate boundaries and a more accurate count of the number of features, but this will not work well if two ground features of the same land cover class are adjacent to one another. For example, a group of trees will be mapped as just one tree if the neighboring pixels or segments are merged together. Another possible solution involves modifying image segment boundaries after classification (e.g. smoothing polygon boundaries, splitting or merging segments) using expert knowledge (i.e. decision rules) and/or additional input data layers (Tiede et al., 2010).

5. Conclusions

In this study, we found that when super-object information was incorporated for supervised classification of a high resolution image of an urban area, overall accuracy was significantly higher than when a single-scale supervised classification approach was used. Overall accuracy for the best classification was 84.42%, and most land cover classes achieved producer's and user's accuracies of 80% or better. We performed classification as super-object information from each of the coarser segmentations was added as variables for classifying the base units (i.e. single pixels or small image segments), and found that overall accuracy increased as super-object information was added up to a certain point (scale parameter of 80), after which it decreased slightly as segments became larger than the actual features of interest. We also tested pixels and image segments as the base units for the classifications that included super-object information, and found that results were best when pixels or small segments were used. When larger segments were used as the base units, classification accuracy decreased due to under-segmentation of small features (e.g. single trees) and features surrounded by spectrally-similar land covers (e.g. buildings surrounded by concrete/asphalt, trees surrounded by grass).

For future studies, it may be interesting to test the use of different feature selection algorithms prior to classification to see if results further improve, and to compare classification accuracy achieved by the random forest algorithm with results obtained using other classification algorithms. Use of additional input data, such as LIDAR height information, may also improve results. Classification methods that incorporate super-object information should also be tested in other types of environments (forested areas, wetlands, etc.) to see if they are applicable in non-urban areas where features have more irregular sizes and shapes. Finally, further research is necessary to identify methods for grouping the classified pixels/image segments into units that more closely approximate features of interest such as individual buildings or trees.

Acknowledgments

We would like to thank the Broward County Property Appraiser's office for providing us with imagery of the study area. We also thank Dr. Caiyun Zhang for her assistance with accuracy assessment.

References

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multiresolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS Journal of Photogrammetry and Remote Sensing 58 (3–4), 239–258.
Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 65 (1), 2–16.
Blaschke, T., Burnett, C., Pekkarinen, A., 2004. New contextual approaches using image segmentation for object-based classification. In: De Meer, F., de Jong, S (Eds.), Remote Sensing Image Analysis: Including the Spatial Domain. Kluwer Academic Publishers, Dordrecht, pp. 211–236.
Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.
Bruzzone, L., Carlin, L., 2006. A multilevel context-based system for classification of very high spatial resolution images. IEEE Transactions on Geoscience and Remote Sensing 44 (9), 2587–2600.
Burnett, C., Blaschke, T., 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. Ecological Modelling 168 (3), 233–249.
Congalton, R., 2009. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. CRC Press, Boca Raton, USA.
Definiens, 2006. Definiens Professional 5 Reference Book. Definiens AG, München, Germany.
Dorren, L., Maier, B., Seijmonsbergen, A., 2003. Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification. Forest Ecology and Management 183 (1–3), 31–46.
Dragut, L., Tiede, D., Levick, S.R., 2010. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. International Journal of Geographical Information Science 24 (6), 859–871.
Gislason, P., Benediktsson, J., Sveinsson, J., 2006. Random forests for land cover classification. Pattern Recognition Letters 27 (4), 294–300.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutmann, P., Witten, I., 2009. The WEKA data mining software: an update. SIGKDD Explorations 11, 1–18.
Ham, J., Chen, Y., Crawford, M., Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 43 (3), 492–501.
Hay, G., Castilla, G., 2006. Object-based image analysis: strengths, weaknesses, opportunities and threats (SWOT). International Archives of Photogrammetry, Remote Sensing and Spatial, Information Sciences XXXVI-4 (C42).
Herold, M., Liu, X., Clarke, K., 2003. Spatial metrics and image texture for mapping urban land use. Photogrammetric Engineering and Remote Sensing 69 (9), 991–1001.
Jensen, J., 2005. Introductory Digital Image Processing: A Remote Sensing Perspective, third ed. Pearson Prentice Hall, Upper Saddle River, USA.
Jin, X., Davis, C., 2005. Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. EURASIP Journal on Applied Signal Processing 14, 2196–2206.
Johnson, B., Xie, Z., 2011. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. ISPRS Journal of Photogrammetry and Remote Sensing 66 (4), 473–483.
Kim, M., Madden, M., Warner, T., 2008. Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery. In: Blaschke, T., Lang, S., Hay, G. (Eds.), Object-based Image Analysis; Spatial Concepts for Knowledge-driven Remote Sensing Applications. Springer, Heidelberg, Berlin, New York, pp. 291–307.
Kim, M., Madden, M., Warner, T., 2009. Forest type mapping using object-specific texture measures from multispectral Ikonos imagery: segmentation quality and image classification issues. Photogrammetric Engineering and Remote Sensing 75 (7), 819–830.
Kim, M., Madden, M., Xu, B., 2010. GEOBIA vegetation mapping in Great Smoky Mountains National park with spectral and non-spectral ancillary information. Photogrammetric Engineering and Remote Sensing 76 (2), 137–149.
Kim, M., Warner, T., Madden, M., Atkinson, D., 2011. Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. International Journal of Remote Sensing 32 (10), 2825–2850.
Lang, S., Schopfer, E., Holbling, D., Blaschke, T., Moeller, M., Jekel, T., Kloyber, E., 2008. Quantifying and qualifying urban green by integrating remote sensing GIS and social science methods. In: Petrosillo et al. (Eds.), Use of Landscape Sciences for the Assessment of Environmental Security. Springer, Netherlands, pp. 93–105.

Lawrence, R., Wood, S., Sheley, R., 2006. Mapping invasive plants using hyperspectral imagery and breiman cutler classifications (RandomForest). Remote Sensing of Environment 100 (3), 356–362.

Liu, D., Xia, F., 2010. Assessing object-based classification: advantages and limitations. Remote Sensing Letters 1 (4), 187–194.

Marpu, P., Neubert, M., Herold, H., Niemeyer, I., 2010. Enhanced evaluation of image segmentation results. Journal of Spatial Science 55 (1), 55–68.

Myint, S., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sensing of Environment 115 (5), 1145–1161.

Openshaw, S., 1984. The modifiable areal unit problem. In: Concepts and Techniques in Modern Geography (CATMOG) 38. Geobooks, United Kingdom. pp. 1–40.

Pal, M., 2005. Random forest classifier for remote sensing classification. International Journal of Remote Sensing 26, 217–222.

Thomas, N., Hendrix, C., Congalton, R., 2003. A comparison of urban mapping methods using high-resolution digital imagery. Photogrammetric Engineering and Remote Sensing 69 (9), 963–972.

Tiede, D., Lang, S., Albrecht, F., Holbling, D., 2010. Object-based class modelling for cadastre-contained delineation of geoobjects. Photogrammetric Engineering and Remote Sensing 76 (2), 193–202.

Trias-sanz, R., Stamon, G., Louchet, J., 2008. Using colour, texture, and hierarchical segmentation for high-resolution remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 63 (2), 156–168.

Walker, J., Blaschke, T., 2008. Object-based land-cover classification for the Phoenix metropolitan area: optimization vs. transportability. International Journal of Remote Sensing 29 (7), 2021–2040.

Walton, J., Nowak, D., Greenfield, E., 2008. Assessing urban forest canopy cover using airborne or satellite imagery. Arboriculture and Urban Forestry 34 (6), 334–340.

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., Schirokauer, D., 2006. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. Photogrammetric Engineering and Remote Sensing 72 (7), 799–811.

Zhou, W., Troy, A., 2009. Development of an object-based framework for classifying and inventorying human-dominated forest ecosystems. International Journal of Remote Sensing 30 (23), 6343–6360.

Zhou, Y., Wang, Y., 2008. Extraction of impervious surface areas from high spatial resolution imagery by multiple agent segmentation and classification. Photogrammetric Engineering and Remote Sensing 74 (7), 857–868.