International Journal of Remote Sensing
Vol. 32, No. 12, 20 June 2011, 3285–3308

Taylor & Francis
Taylor & Francis Group

Object-based urban detailed land cover classification with high spatial resolution IKONOS imagery

RUILIANG PU*†, SHAWN LANDRY‡ and QIAN YU§
†Department of Geography, University of South Florida, 4202 E. Fowler Avenue, NES 107, Tampa, FL 33620, USA
‡Florida Center for Community Design and Research, 4202 E. Fowler Avenue, HMS 301, Tampa, FL 33620, USA
§Department of Geosciences, University of Massachusetts, 611 N. Pleasant St, Amherst, MA 01003, USA

(*Received 7 November 2009; in final form 8 February 2010*)

Improvement in remote sensing techniques in spatial/spectral resolution strengthens their applicability for urban environmental study. Unfortunately, high spatial resolution imagery also increases internal variability in land cover units and can cause a 'salt-and-pepper' effect, resulting in decreased accuracy using pixel-based classification results. Region-based classification techniques, using an image object (IO) rather than a pixel as a classification unit, appear to hold promise as a method for overcoming this problem. Using IKONOS high spatial resolution imagery, we examined whether the IO technique could significantly improve classification accuracy compared to the pixel-based method when applied to urban land cover mapping in Tampa Bay, FL, USA. We further compared the performance of an artificial neural network (ANN) and a minimum distance classifier (MDC) in urban detailed land cover classification and evaluated whether the classification accuracy was affected by the number of extracted IO features. Our analysis methods included IKONOS image data calibration, data fusion with the pansharpening (PS) process, Hue–Intensity–Saturation (HIS) transferred indices and textural feature extraction, and feature selection using a stepwise discriminant analysis (SDA). The classification results were evaluated with visually interpreted data from high-resolution (0.3 m) digital aerial photographs. Our results indicate a statistically significant difference in classification accuracy between pixel- and object-based techniques; ANN outperforms MDC as an object-based classifier; and the use of more features (27 vs. 9 features) increases the IO classification accuracy, although the increase is statistically significant for the MDC but not for the ANN.

1. Introduction

Timely and accurate information on the status and trends of urban land cover and biophysical parameters is crucial when developing strategies for sustainable development and improving urban residential environmental and living quality (Yang *et al.* 2003, Song 2005). Developing techniques that enhance our ability to monitor and map urban land cover are therefore important for city planning and management.

*Corresponding author. Email: rpu@usf.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2011 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/01431161003745657

One of the most common applications of remote sensing images is the extraction of land cover information for digital image base maps. Such information is useful to city governments seeking better planning and management approaches to deal with the numerous problems associated with increasing urbanization (e.g. urban heat island and traffic congestion). During the past decade, advances in satellite remote sensing have included finer spatial (e.g. IKONOS multispectral (MS) images at 4-m resolution and panchromatic band at 1-m resolution) and spectral resolution (e.g. Hyperion hyperspectral sensor at 10-nm spectral resolution). The use of high spatial resolution commercial satellite imagery (e.g. IKONOS) has been shown to be cost-competitive with traditional aerial photographic surveys for generating digital image base maps (Davis and Wang 2003). However, the use of high spatial resolution imagery also poses challenges for urban land cover classification.

Local variance in different environments within an image scene is a function of the sizes and spatial relationships of the land cover classes. The relatively high local variance of urban land cover classes presents challenges for automatic classification using high spatial resolution satellite sensors (Woodcock and Strahler 1987). Because of the improvement in spatial resolution, land cover classes tend to be represented by spatial units of heterogeneous spectral reflectance characteristics and their statistical separability is limited using traditional pixel-based classification approaches. Consequently, classification accuracy is reduced and the results usually show a 'salt-and-pepper' effect when individual pixels are classified differently from their neighbours. Studies have shown a decrease in land cover classification accuracy associated with an improvement in image spatial resolution, when other sensor characteristics are kept unchanged (Townshend and Justice 1981, Latty *et al.* 1985, Martin *et al.* 1988, Gong and Howarth 1990, Treitz and Howarth 2000). Classification accuracy is particularly problematic in urban environments, which typically consist of mosaics of small features made up of materials with different physical properties (Mathieu *et al.* 2007). To overcome this problem, region- or object-based classification can be used. Object-based techniques first use image segmentation to produce discrete regions or image objects (IOs) that are more homogeneous in themselves than with nearby regions, and then use these IOs rather than pixels as the unit for classification (Carleer and Wolff 2006, Blaschke 2010).

An object-based classification strategy can potentially improve classification accuracy compared to pixel-based classification for several reasons: (1) partitioning an image into IOs is similar to the way humans conceptually organize the landscape to comprehend it (Hay and Castilla 2008); (2) in addition to spectral features, IOs enable the use of texture and contextual (relationships with other objects) features and some shape/geometric features (e.g. form, size and geomorphology) (Yu *et al.* 2006, Hay and Castilla 2008); and (3) the objects of interest to be extracted from a certain scene can be associated with different abstraction levels (i.e. different scales) and these levels can be represented in an analysis system (Kux and Araújo 2008). Previous researchers have demonstrated the advantages of object-based classification (Ton *et al.* 1991, Johnsson 1994, Hill 1999, Herold *et al.* 2003, Carleer and Wolff 2006, Kong *et al.* 2006, Marchesi *et al.* 2006, Yu *et al.* 2006, Mathieu *et al.* 2007, Kux and Araújo 2008). For example, Kong *et al.* (2006) adopted an object-based image analysis approach using multifeature information of the IOs and multiscale image segmentation technology to classify urban land cover information (urban roads, buildings, woods, farmland and waters) from high-resolution imagery (2.4-m spatial resolution and four MS bands) in Wuhan City, China. Their experimental results indicated that the object-based analysis approach offered a satisfactory solution for extracting the required information rapidly and efficiently. Yu *et al.* (2006) used an object-based detailed vegetation classification with the Digital Airborne Imaging System (DAIS; four MS bands with 1-m spatial resolution), an airborne high spatial resolution remote sensing system, at Point Rayes National Seashore Park in northern California, USA. They demonstrated that the object-based classification overcame the problem of the salt-and-pepper effects found in the classification results from traditional pixel-based approaches and thus improved the classification accuracy. Mathieu *et al.* (2007) used an object-based classification to map large-scale vegetation communities with IKONOS imagery in Dunedin City, New Zealand. Although the approach did not provide maps as detailed as those produced by manual interpretation of aerial photographs, they found it was possible using the object-based approach to extract ecologically significant classes and it was an efficient way to generate accurate and detailed maps.

Our literature review indicates that more work is needed to evaluate object-based classification approaches with 1–5 m high-resolution imagery, especially the efficiency of such approaches with regard to urban detailed land cover classification. The following problems are addressed in this paper. First, given a relatively high local variance of urban land cover classes, most existing efforts related to mapping land cover in an urban environment have focused on mapping several coarse land cover types (e.g. Kong *et al.* 2006, Marchesi *et al.* 2006), and have not adequately used the detailed spatial information that high spatial resolution imagery can provide for a detailed urban land cover classification. Second, although some researchers have tested rule-based classifiers (e.g. Santos *et al.* 2006), most studies have used either the simple classifier (nearest-neighbour classifier) provided by eCognition software or a traditional parametric classifier such as the maximum likelihood classifier (MLC) or the minimum distance classifier (MDC) (e.g. Carleer and Wolff 2006, Marchesi *et al.* 2006, Yu *et al.* 2006, Kux and Araújo 2008). In this study, we evaluate the performance of the non-parametric artificial neural network (ANN; Gong *et al.* 1997) classifier in addition to the traditional MDC for urban detailed land cover classification using high-resolution image data. The ANN was tested with both pixel- and object-based classification units, while MDC was tested with object-based data only. The ANN was selected as a classifier because it is a non-parametric and non-linear algorithm. ANN is expected to make efficient use of subtle spectral differences in the MS images due to ANN's multilayer structure, to handle data at any measurement scale (e.g. some features extracted from IOs) and to use a relatively small sample size for training purposes for classification. Finally, most existing studies have not used a method to select a subset of features from a set of candidate features extracted from and characterizing IOs. Selecting a subset from all possible features extracted from IOs is necessary because of the information redundancy that exists among the features, including spectral, textual/contextual and shape/geometric features.

Therefore, the objectives of this study were: (1) to test whether the object-based technique could significantly improve classification accuracy compared to a pixel-based method when applied to urban detailed land cover classification using high spatial resolution imagery; (2) to compare the performance of ANN and MDC for urban detailed land classification using high-resolution data; and (3) to evaluate whether urban land cover classification accuracy is affected by the number of extracted IO features. Some limitations associated with the use of the object-based classification approach are also discussed.

2. Study area and data sets

2.1 *Study area*

The study area is a 100 km^2 area within the City of Tampa. Tampa is the largest city on the west coast of Florida and consists of approximately 285 km^2. During the past three decades, the city had experienced continuous growth in population and expansion in its extent. The population is currently estimated at approximately 335 000 (www.tampagov.net, accessed on 26 November 2007). The city is located at approximately 28°N and 82°W (figure 1). Historically, the natural plant communities of the Tampa Bay region included pine flatwoods, cypress domes, hardwood hammocks, high pine forests, freshwater marshes and mangrove forests. Based on the City of Tampa Urban Ecological Analysis (Andreu *et al*. 2008), tree canopy cover is 28.1%. Other green vegetation areas are occupied by shrubs, grass/lawns of various sizes, golf courses and crops. Man-made materials for buildings and roofs in the city include concrete, metal plate, brick tile, etc. Various impervious road surfaces are covered by asphalt, concrete and rail track.

2.2 *Data sets*

Three data sets were used in this study: IKONOS imagery, digital aerial photographs and ground plot measurements. (1) High-resolution IKONOS satellite imagery (GeoEye, Inc., Virginia, USA) was acquired for the study area on 6 April 2006. Georeferenced 1-m resolution panchromatic (Pan, 526–929 nm) and 4-m resolution MS images (four bands: blue (445–516 nm), green (506–595 nm), red (632–698 nm)

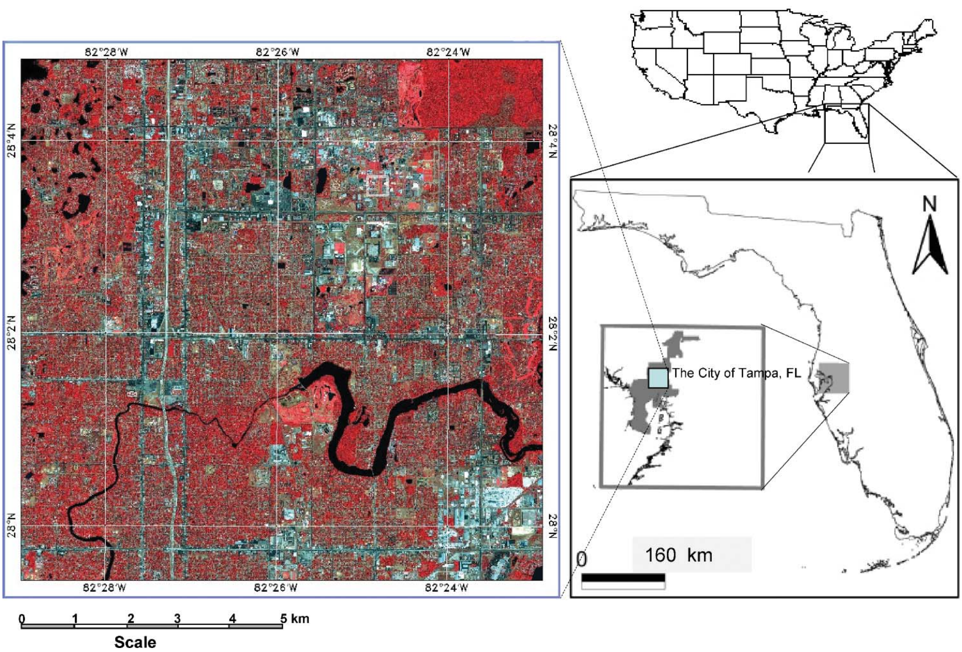


Figure 1. A location map of the study area. The grey box in the image shows the location of the portion of land cover classification results in figure 3.

and near-infrared (NIR, 757–853 nm) were acquired. The IKONOS imagery including Pan and MS images was the primary data set for this object-based classification analysis. (2) A set of true colour digital aerial photographs was taken in January 2006 (SWFWMD 2006). The aerial photographs included three visible bands (blue, green and red) at 0.3-m spatial resolution. They were used as reference to define training, test and validation areas/samples. (3) Measurements from approximately 60 ground plots in the study area, including land cover type and percentage, plant species, diameter at breast height (dbh) and crown width, were provided by the City of Tampa Ecological Analysis 2006–2007 (Andreu *et al*. 2008). Ground plot measurements were used as reference for determining training and test areas.

3. Methodology

Figure 2 presents a flowchart of the analysis procedure, which included urban land cover classification using high-resolution IKONOS imagery with both pixel- and object-based classification strategies. Preprocessing of IKONOS imagery data included radiometric correction and calibration and data fusion. Nine basic pixel-based image layers were prepared, including four pansharpening (PS) bands, three Hue–Intensity–Saturation (HIS) indices, one soil-adjusted vegetation index (SAVI), and one texture image (created from PS band 4 with co-occurrence and homogeneity parameters from ENVI (ITT 2006)). The selection of SAVI was considered to compress the effect of bare soil and other non-vegetated background on vegetation spectra; the extraction of three HIS indices was for adequately using spectral information contained in the three IKONOS visible bands while the texture image layer

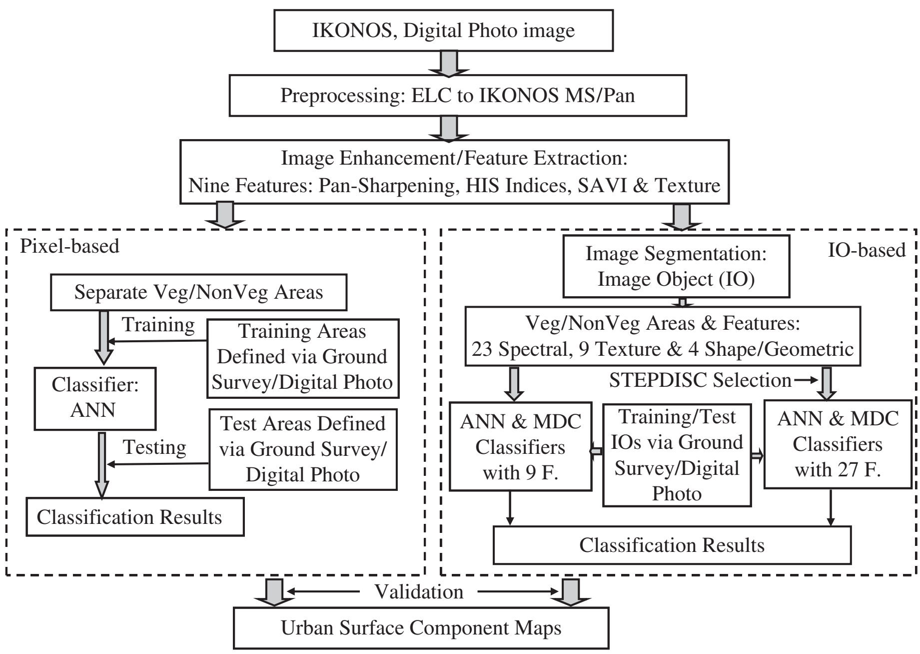


Figure 2. A flowchart of the urban land cover mapping analysis procedure, consisting of pixel- and object-based classification strategies. Veg, vegetated; NonVeg, non-vegetated; STEPDISC, an SAS procedure that was used to select important features (F).

was expected to help separate different vegetation types (e.g. grass/lawn and tree canopy). All of the nine pixel-based image layers were rescaled to [0, 10 000]. They were then used for the pixel-based classification and for creating IOs to test the object-based classification. In addition to the IOs generated from the nine image layers (themselves forming nine features), 27 more features were extracted from the IOs and used for object-based classification analysis.

3.1 *Data preprocessing and fusion*

The four MS IKONOS bands and one Pan band were radiometrically corrected to ground reflectance using the Empirical Line Calibration (ELC) outlined by Jensen (2005) and the calibration parameters provided by Space Imaging in the imagery metadata. The four MS IKONOS image bands were calibrated using the ELC method to convert the radiance to ground surface reflectance using ENVI software. The ground spectral measurements were taken using an ASD spectrometer (FieldSpec®3, Analytical Spectral Devices, Inc., Colorado, USA) from river (deep/clear water) and sand targets located within the IKONOS image area. To calibrate the Pan image, the IKONOS Pan relative spectral response was first resampled to a 50-nm wavelength interval and then the average reflectance for the Pan bandwidth was calculated from the ASD spectral measurements. Consequently, the four MS bands and one Pan band were calibrated to ground reflectance prior to data fusion processing.

Pansharpened 1-m resolution images were created by fusing the 4-m MS IKONOS image with the 1-m Pan IKONOS imagery using the method of Principal Component Spectral Sharpening (ITT 2006). Although PS altered the spectral values of the images and may have created spectral artefacts that impacted accuracy measures, the use of PS imagery has been shown to improve the classification of forests (Kosaka *et al.* 2005) and has been used successfully in urban areas (Jain and Jain 2006, Nichol and Wong 2007) because the pansharpened image reflects the synergic effectiveness of MS and Pan images. A visual comparison of the original colour composite image (4-m resolution) with the pansharpened colour composite image (nominal 1-m resolution) revealed that urban land cover boundaries on the latter image were clearer and smoother than those on the former image, especially for some tree crown boundaries and relatively small IOs. Therefore, all further image analysis including image segmentation and feature extraction used the four PS band imagery.

3.2 *Image segmentation*

The object-based image analysis software used in this research was Definiens eCognition 5.0. eCognition uses a multiresolution segmentation approach that is a bottom-up region-merging technique starting with one-pixel objects. In numerous iterative steps, smaller IOs are merged into larger ones (Baatz *et al.* 2004). The merging criterion minimizes the average heterogeneity of IOs weighted by their size in pixels (Baatz and Schape 2000, Benz *et al.* 2004). Quantitatively, the definition of heterogeneity takes into account both the spectral variance and geometry of the objects (Yu *et al.* 2006). The outcome of the segmentation algorithm is controlled by a scale factor and a heterogeneity criterion. The heterogeneity criterion controls the merging decision process and is computed using spectral layers (e.g. MS images) or non-spectral layers (e.g. thematic data such as elevation) (Mathieu *et al.* 2007). The heterogeneity criterion includes two mutually exclusive properties: colour and shape.

Colour refers to the spectral homogeneity whereas shape considers the geometric/geomorphologic characteristics of the objects. Shape is further divided into two equally exclusive properties: smoothness and compactness (Baatz *et al.* 2004).

The optimum segmentation parameters depend on the scale and nature of the features to be detected. Previous researchers (Mathieu *et al.* 2007) have used a systematic trial-and-error approach validated by the visual inspection of the quality of the output IOs (i.e. how well the IOs matched feature boundaries in the image for a particular application). Once an appropriate scale factor was identified, the colour and shape criteria were modified to refine the shape of the IOs. Most previous studies had found that more meaningful objects were extracted with a higher weight for the colour criterion (e.g. Laliberte *et al.* 2004, Mathieu *et al.* 2007). Using the input of nine data layers (four PS bands, three HIS indices, one SAVI and one texture image) for urban detailed land cover mapping, the colour criterion was assigned a weight of 0.7, whereas the shape received the remaining weight of 0.3. Furthermore, the compactness was assigned a weight of 0.3 and the smoothness was assigned the remaining weight of 0.7. Three different scales (70, 100 and 150) of image segmentation were tested. After visually inspecting the degree to which IOs matched the feature boundaries of the land cover types in the study area, we used the IOs created with a scale of 70 in the remainder of the object-based classification analysis.

3.3 *Feature extraction and selection*

In addition to the nine features (eight spectral features and one texture feature) used for creating the IOs, 27 more feature variables were generated from each IO. Table 1 lists all 36 features (23 spectral features, nine texture features and four shape/geometric features) used for the object-based classification analysis. The inclusion of these features was based on previous studies (e.g. Haralick *et al.* 1973, Carleer and Wolff 2006, Kong *et al.* 2006, Yu *et al.* 2006).

Table 1 shows 23 spectral features consisting of means and standard deviations of four PS bands, three HIS transfer indices, one SAVI and one textural (as input pixel-based data layers) and four ratios of four PS band features and one brightness feature calculated by averaging means of the four PS bands. The four ratio features were calculated by the band *i* mean value of an IO divided by the sum of all four PS bands. The nine texture features include five grey-level co-occurrence matrix (GLCM) textures and four grey-level difference vector (GLDV) textures calculated from PS band 4 (similar to the NIR band). The GLCM indicates the frequency at which different combinations of grey levels of two pixels at a fixed relative position occur in an IO. A different co-occurrence matrix exists for each spatial relationship. The GLDV is the sum of the diagonals of the GLCM and counts the occurrence of references to the neighbouring pixels' absolute differences. Compared to pixel-based texture, which is generally calculated using a specific window size, the GLCM and GLDV calculate the texture for all pixels of an IO (Haralick *et al.* 1973, Yu *et al.* 2006). The remaining four shape/spatial features are described by two compactness and two shape/geometric indices.

To reduce redundancy, it was necessary to select a subset of features from the 36 feature variables prior to the object-based classification. In this analysis we used stepwise discriminant analysis (SDA) to select features based on minimizing within-class variance while maximizing the between-class variance for a given significance level. This was a relatively effective method to select a subset of the quantitative

Table 1. Image-object (IO) features used in this analysis.

Feature name	Description
Band1	Mean of pansharpened IKONOS band1 (blue), input pixel layer
Band2	Mean of pansharpened IKONOS band2 (green), input pixel layer
Band3	Mean of pansharpened IKONOS band3 (red), input pixel layer
Band4	Mean of pansharpened IKONOS band4 (NIR), input pixel layer
Hue	Mean of Hue image processed from pansharpened IKONOS bands 3, 2, 1, input layer
Sat	Mean of Saturation image processed from pansharpening IKONOS bands 3, 2, 1, input layer
Val	Mean of Value (Intensity) image processed from pansharpened IKONOS bands 3, 2, 1, input layer
SAVI	Mean of soil-adjusted vegetation index (SAVI): 1.5 (band4 − band3)/(band4 + band3 + 0.5), input layer
Tex	Mean of texture information of co-occurrence homogeneity extracted from band4, input layer
SDB1	Standard deviation of band1
SDB2	Standard deviation of band2
SDB3	Standard deviation of band3
SDB4	Standard deviation of band4
SDH	Standard deviation of Hue
SDS	Standard deviation of Sat
SDV	Standard deviation of Val
SDVI	Standard deviation of SAVI
SDTX	Standard deviation of Tex
Ratio1	Band1 mean divided by sum of band1 to band4 means
Ratio2	Band2 mean divided by sum of band1 to band4 means
Ratio3	Band3 mean divided by sum of band1 to band4 means
Ratio4	Band4 mean divided by sum of band1 to band4 means
Bright	Brightness, average of means of bands 1 through 4
GLCMH	GLCM homogeneity from band4, $\sum_{i,j=0}^{N-1} \frac{p_{i,j}}{1+(i-j)^2}$
GLCMCON	GLCM contrast from band4, $\sum_{i,j=0}^{N-1} p_{i,j}(i-j)^2$
GLCMD	GLCM dissimilarity from band4, $\sum_{i,j=0}^{N-1} p_{i,j}\lvert i-j\rvert$
GLCME	GLCM entropy from band4, $\sum_{i,j=0}^{N-1} p_{i,j}(-\ln p_{i,j})$
GLCMSD	GLCM standard deviation from band4, $\sigma_{i,j}^2 - \sum_{k=0}^{N-1} p_{i,j}(i, j-\mu_{i,j})$ where $\mu_{\mathrm{i,j}} = \sum_{\mathrm{k=0}}^{\mathrm{N-1}} \mathrm{p_{i,j}}/\mathrm{N}^2$
GLCMCOR	GLCM correlation from band4, $\sum_{i,j=0}^{N-1} p_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right]$
GLDVA	GLDV angular second moment from band4, $\sum_{k=0}^{N-1} V_k^2$
GLDVE	GLDV entropy from band4, $\sum_{k=0}^{N-1} V_k(-\ln V_k)$
GLDVC	GLDV contrast from band4, $\sum_{k=0}^{N-1} V_k K^2$
Compact	Compactness, the product of the length and the width of the corresponding object and divided by the number of its inner pixels
CompactP	Compactness, the ratio of the area of a polygon to the area of a circle with the same perimeter
ShapeI	Shape index, the border length of the IO divided four times the square root of its area, i.e. smoothness.
NumP	Number of edges, the number of edge that from the polygon.

i, row number; *j*, column number; $V_{i,j}$, the value in the cell i, j of the matrix; $p_{i,j}$, the normalized value in the cell i, j; N, the number of rows or columns.

variables for use in discriminating among the classes and has been applied by many researchers for reduction of data dimensionality (e.g. Clark *et al.* 2005, van Aardt and Wynne 2007). In this study, we used the SAS STEPDISC Procedure (SAS 1991) with training samples to select a subset of features from the 36 feature variables with a p-value <0.001.

3.4 *Classification strategies*

To improve the urban land cover classification accuracy, a hierarchical classification system constructed using three levels was adopted for the study, which matched the logical structure of most land cover classification schemes used by previous researchers (e.g. Townsend and Walsh 2001, Pu *et al.* 2008). The hierarchical classification scheme, including land cover classes and descriptions, is presented in table 2. At level I, vegetated and non-vegetated areas were separated using an SAVI threshold of 0.19, where values greater than 0.19 were assigned as vegetation. The threshold of 0.19 was determined by first checking the histogram of the SAVI image to find an approximate threshold (e.g. 0.2), then adjusting the threshold slightly based on visual inspection to separate vegetated and non-vegetated areas. At level II, the vegetated and non-vegetated areas were further subdivided into five vegetated and four non-vegetated classes. The five vegetated types included broadleaf trees (BT), needleleaf trees (NT), palm trees (PT), shrub (Sh) and grass/lawn (GL). The four non-vegetated classes included building/roof (BR), impervious area (IA), sand/soil (SS) and water (Wa). At level III, one vegetated class, BT, was further subdivided into high NIR reflectance (BT1) and low NIR reflectance (BT2). This separation of BT was chosen in consideration of the significant difference of the NIR reflectance between sand live oak and most other BT species, presumably a result of biological characteristics (e.g. deciduous vs. even green). The two non-vegetated classes, BR and IA, were further subdivided into high, medium and low albedo (BR1, BR2 and BR3; IA1, IA2 and IA3), respectively. Classification operations were carried out at level III separately for each level I area (vegetated/non-vegetated) using pixel- or object-based features with ANN and MDC algorithms (figure 2). The final classification results at level II were obtained by merging BT1 and BT2 into BT, BR1 to BR3 into BR, and IA1 to IA3 into IA. All the accuracy indices were calculated at level II.

Two supervised classification algorithms were used for the urban land cover classification: the non-parametric ANN and the parametric MDC. In this analysis, a feedforward ANN algorithm was used for classifying the 14 level III classes. The network training mechanism was an error-propagation algorithm (Rumelhart *et al.* 1986, Pao 1989). A neural network program developed by Pao (1989) was adapted for use in this study. In a layered structure, the input to each node is the sum of the weighted outputs of the nodes in the prior layer, except for the nodes in the input layer, which are connected to the feature variables (as discussed in section 4). The nodes in the last layer output a vector corresponding to the 14 level III classes. Layers between input and output layers are called hidden layers. One hidden layer ($h1$) has been shown to be sufficient for most learning purposes (Gong *et al.* 1997). The learning procedure is controlled by three variables: a learning rate (η, defined as the network's learning speed of reducing system error, usually taking any value from a range [0, 1]); a momentum coefficient (α, a new weight change is modified by including some (i.e. α) of the weight change in the previous iteration of network learning in order to weaken system oscillations, and usually it also takes any value from a range [0, 1]);

Table 2. Urban land cover classes, definitions, number of training/test samples (IOs and pixels) and evaluation points used in this analysis.

	Level II			Level III				
Level 1	Name	Abbreviation	Description	Abbreviation	Description	No. of training/test IOs	No. of training/test pixels	Evaluation points
Vegetated area	Broadleaf trees	BT	All broadleaf tree species canopies	BT1	High NIR reflectance	171	10 738	109
				BT2	Low NIR reflectance	169	13 822	50
	Needleleaf trees	NT	All conifer tree species canopies	NT	–	82	3 926	0
	Palm trees	PT	All palm tree species canopies	PT	–	71	2 000	2
	Shrub	Sh	All shrub, bush, including some bush in wetland, pond and lake side.	Sh	–	86	6 770	20
	Grass/lawn	GL	All grassland, golf course and lawns	GL	–	90	7 766	46
Non-vegetated area	Building/roof	BR	All different sized buildings or roofs covered with different materials	BR1	High albedo	181	6 191	15
				BR2	Medium albedo	147	8 309	33
				BR3	Low albedo	143	8 292	12
	Impervious area	IA	All impervious surface areas, e.g. roads, parking lots	IA1	High albedo	135	7 810	36
				IA2	Medium albedo	148	13 763	51
				IA3	Low albedo	143	12 848	13
	Sand/soil	SS	All bare sand/soil and/or very dry/dead grasslands.	SS	–	99	10 209	37
	Water	Wa	All different types of water bodies	Wa	–	73	15 106	17
Total of training/test samples						1 738	127 550	441

and a number of nodes in one hidden layer ($h1$). All the three variables need to be specified empirically based on the results of a limited number of tests. The network training is done by repeatedly presenting training samples (pixel or IO samples) with a known class and terminates when the output meets a minimum error criterion or optimal test accuracy is achieved. In this study, we used an optimal test accuracy calculated from test samples to control the network training processing. The trained network was then used to classify the 14 level III classes separately using the nine or 27 selected feature variables.

For comparison with the result from the ANN algorithm, we conducted additional object-based classifications of the 14 level III classes using the MDC with input of nine or 27 selected feature variables. The SAS DISCRIM procedure with Mahalanobis distance to determine proximity (SAS 1991) was used for the MDC classification. This method was chosen because the parametric algorithm was expected to yield good results even when the assumed class distribution (normal distribution) was invalid for some features of IOs (Schowengerdt 2007). For the object-based classification, nine features (corresponding to the nine input data layers) and 27 features (selected from a total of 36 features by using the SDA procedure) were used with the ANN and MDC algorithms.

3.5 *Training, test and evaluation samples*

Training and test samples used for the classification algorithms and comparative accuracy assessments were determined from pixel- and object-based image data by referencing 0.3-m resolution digital aerial photographs and measurements from approximately 60 ground plots. Table 2 summarizes the number of training and test samples corresponding to each of the 14 level III classes for the object- and pixel-based classifications. The total number of IO samples used for training and testing for the object-based classification was 1738, extracted from the segmented image at scale = 70. The total number of pixels used for training and testing for the pixel-based classification was 127 550 with a minimum of 2000 pixels for each class.

Pixel-based training/test samples were defined as regions of interest (ROIs) using the ENVI function of defining training areas (ITT 2006) on an IKONOS image by referring to 0.3-m aerial photograph and available ground plot data. When locating the training and test samples, we considered the ROIs' representativeness and randomness corresponding to the 14 level III classes in the study area (i.e. each ROI contained multiple patches representing each class). Three separate ROIs were created for each class and used to develop three different sets of pixel-based training and test datasets, where each set contained two regions for training and one region for testing. The object-based training and test samples were determined from the segmented image (IO image) overlaid with the ground plot location layer (i.e. the dataset from Andreu *et al.* 2008) and were selected by referencing the 0.3-m resolution aerial photographs and ground plot data (i.e. land cover type/percentage and plant species). Only IOs that had coincident boundaries with real ground objects as seen in the aerial photograph were selected as training/test samples. With a systematic sampling approach, about two-thirds of the selected IO samples were used for training and about one-third for testing. This procedure was repeated three times (runs) to obtain three different sets of test samples (but training sets with a part overlaid between any two training sets). Although this sampling strategy has been shown to overestimate accuracy due to spatial autocorrelation of neighbouring pixel or IO samples (Muchoney and Strahler 2002), the technique was considered acceptable for a comparative accuracy assessment. Comparative accuracy assessments were calculated from ANN and MDC level II classification results using a confusion matrix constructed with the test samples. Comparative accuracy measures included average accuracy (AA, defined as the average producer's accuracy), overall accuracy (OAA, defined as the ratio of correct classified samples to total samples) and the kappa index (Congalton and Mead 1983, Congalton and Green 1999).

In addition to the comparative accuracy assessment described above, we used a system sampling approach with a 500-m grid to validate the land cover classification results. A total of 441 points (cross-points of the 500-m grids, labelled evaluation points in table 2) each representing about 4 m^2 were visually identified and interpreted from both the digital aerial photographs and resultant urban land cover classification maps. As individual NT and PT and their spatial distribution were too small in the study area, neither NT nor PT points were available for evaluating classification accuracy. As a result of this limitation in evaluation points, we used test samples to verify the performance of different classifiers and effectiveness of classification units. It was expected that both test/evaluation samples would produce reasonably similar results for the purpose of evaluating performance. An OAA value and a kappa index were calculated from the 441 paired-points and used for a general assessment of accuracy of the land cover classification maps produced using either pixel- or object-based IKONOS image data with the ANN and MDC algorithms.

4. Results and analysis

4.1 *Feature selection*

An SDA procedure was first performed to select a subset of important feature variables to use with the object-based classification. Table 3 shows the 27 features selected using the stepwise procedure. Features with a p-value <0.001 were selected, while those features after step 27 had a p-value >0.001 and thus were eliminated from this analysis. Among the 27 selected feature variables, 19 were spectral features. Selection of all spectral features in the first eight steps indicated that those spectral features made a substantial contribution to separating most of the 14 level III classes. The nine removed feature variables (i.e. p-value >0.001) included four spectral features, three textural features and two shape/geometric features. Two of the nine removed features were shape/geometric features related to compactness, from which it could be interpreted that 'compactness' of IOs had a relatively low separability among the 14 level III classes. Five of the removed features were related to 'contrast and standard deviation' of spectral and textural features (SDB1, SDV, GLCMCON, GLCMD and GLDVC; see table 1 for definitions), which indicated that the 'contrast and standard deviation' information of the five features within an IO was not significantly variable across the 14 classes. The remaining two features removed, Band2 and Ratio4, may have been removed simply because the information they provided was redundant compared with the information provided by the selected variables. From the selected spectral features, it was obvious that the ability to separate the 14 classes relied mainly on the variation of pixel spectral information that was extracted from and characterized the IOs.

Table 3. Summary of 27 features selected using the stepwise discriminant analysis (SDA) procedure.

Step	Feature*	Partial R^2	F-value	Pr > F
1	SAVI	0.8411	701.74	<0.0001
2	Band3	0.7318	361.71	<0.0001
3	Bright	0.5514	162.79	<0.0001
4	Ratio2	0.5257	146.72	<0.0001
5	Band4	0.4411	104.43	<0.0001
6	Ratio3	0.2818	51.88	<0.0001
7	SDB3	0.2367	40.98	<0.0001
8	Band1	0.1497	23.25	<0.0001
9	GLCME	0.1457	22.52	<0.0001
10	Sat	0.1153	17.20	<0.0001
11	SDVI	0.0984	14.39	<0.0001
12	GLCMH	0.0901	13.06	<0.0001
13	Ratio1	0.1272	19.20	<0.0001
14	Hue	0.0711	10.07	<0.0001
15	Val	0.0628	8.81	<0.0001
16	Shapel	0.0635	8.91	<0.0001
17	SDH	0.0585	8.17	<0.0001
18	Tex	0.0537	7.45	<0.0001
19	SDB2	0.0494	6.82	<0.0001
20	SDTX	0.0375	5.11	<0.0001
21	NumP	0.0350	4.76	<0.0001
22	GLCMCOR	0.0291	3.93	<0.0001
23	GLCMSD	0.0296	3.99	<0.0001
24	GLDVA	0.0293	3.95	<0.0001
25	GLDVE	0.0302	4.08	<0.0001
26	SDB4	0.0280	3.77	<0.0001
27	SDS	0.0239	3.19	<0.0001

*See table 1 for full names of the features.

4.2 *Pixel-based classification*

The pixel-based classification results were produced using the ANN algorithm with the 14 ROIs delineated as training areas from PS bands 4, 3 and 2 and using the 0.3-m digital aerial photographs and available ground plot measurements as reference. As stated previously, six vegetated classes and eight non-vegetated classes (i.e. level III classes) were first separately classified and then merged to nine level II classes. Based on the high AA and kappa value calculated from the test samples, a final set of ideal structure parameters of ANN for pixel-based classification with nine features was adopted (learning rate (η) = 0.2, momentum coefficient (α) = 0.8 and number of nodes in a hidden layer ($h1$) = 12 or 10). The results of the pixel-based classification using ANN are presented in figure 3(a) and reveal the 'salt-and-pepper' phenomenon usually caused by using high spatial resolution and pixel-based image data. Table 4 lists the corresponding classification accuracy indices: AA, OAA and kappa values. All the results shown in table 4 were calculated by averaging the three sets of results produced from the three sets of test samples.

4.3 *Object-based classification*

The better ANN structure parameters for the object-based classification with inputs of either nine or 27 features were found by testing various combinations of η, α and $h1$

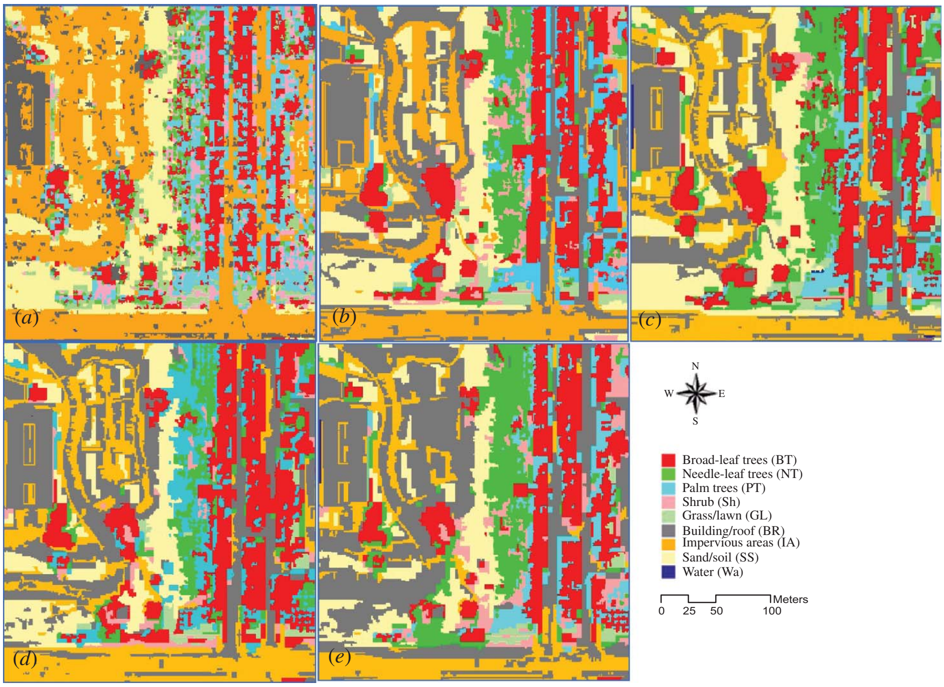


Figure 3. Classification results of urban land cover classes, showing a small portion (figure 1) of the study area at 1:1 scale: (*a*) using nine features of the pixel-based IKONOS imagery with the ANN algorithm; (*b*) using 27 and (*c*) nine features of the object-based IKONOS imagery with the ANN algorithm; and (*d*) using 27 and (*e*) nine features of the object-based IKONOS imagery with the MDC algorithm.

Table 4. Accuracy of urban detailed land cover classification using different classification units (pixel- and object-based) and different algorithms (ANN and MDC) with nine or 27 features (bands). The accuracy in a cell is the average of accuracies calculated from the three sets of test samples.

		Accuracy (%)				Kappa value			
		Pixel-based		IO-based		Pixel-based		IO-based	
Algorithm	No. of features	AA	OAA	AA	OAA	Kappa value	Variance	Kappa value	Variance
ANN	9	73.58	73.82	76.69	78.48	0.6956	0.000030	0.7371	0.000454
	27	n/a	n/a	80.75	81.53	n/a	n/a	0.7736	0.000410
MDC	9	n/a	n/a	68.16	71.12	n/a	n/a	0.6474	0.000561
	27	n/a	n/a	75.34	78.02	n/a	n/a	0.7296	0.000476

AA, average accuracy; OAA, overall average.

using the first training/test data set. For the input of nine features, the better ANN structure parameters were $\eta = 0.8$ or 0.7, $\alpha = 0.2$ or 0.1 and $h1 = 15$ or 12. For the input of 27 features, the better ANN structure parameters were: $\eta = 0.6$ or 0.8, $\alpha = 0.2$ or 0.3 and $h1 = 20$ or 25. Figure 3(*b*) shows the object-based classification result using the ANN algorithm and the 27 input features. Figure 3(*c*) shows the results using the nine input features. Visual inspection of the 1:1 scale maps reveals that the classification result created with the 27-feature input was better than that with the nine-feature, especially for the BT and NT vegetation classes. Table 5 presents the two confusion matrices and three accuracy indices (i.e. AA, OAA and kappa) created with (A) 27 features and (B) nine features. Both matrices were produced with the second set of test samples, which resulted in median accuracy (because both the 27- and nine-input features had three sets of test samples). A comparison of the two sets of results in table 5 shows that AA, OAA and kappa accuracy indices using the 27-feature input are higher (i.e. more accurate) than those using the nine-feature input, although the producer's accuracy is higher for some individual classes of the nine-feature input. The results in table 4 also indicate that using more feature variables produced better results than using fewer feature variables.

We also compared the object-based classification results produced using the ANN algorithm with those produced using MDC, a traditional classifier. Figures 3(*d*) and 3(*e*) show the classification results with inputs of 27 or nine features using the MDC algorithm, respectively. Comparison of the MDC (figures 3(*d*) and 3(*e*)) with the ANN (figures 3(*b*) and 3(*c*)) results using 1:1 scale maps reveals too much PT area on figure 3(*d*) (MDC) compared with that in figure 3(*b*) (ANN) using the 27 input features, and too much BR area in figure 3(*e*) (MDC) compared with that in figure 3(*c*) (ANN) using the nine input features. The results in table 4 indicate that the accuracy indices are much lower for the object-based classification using the MDC algorithm compared to the ANN algorithm with either nine or 27 input features. These results demonstrate that the ANN algorithm outperforms the MDC when using the same number of input features.

4.4 *Evaluation*

The accuracy of the land cover classification maps was evaluated using the 441 evaluation points. An OAA value, the kappa value and the producer's and user's accuracies (Story and Congalton 1986) were calculated for all of the classification maps (table 6). Because of the lack of NT identified at the grid points (possibly caused by the low frequency of NT in the study area), AA could not be calculated for the evaluation. Table 6 presents the evaluation results (OAA and kappa values) produced by the ANN algorithm with nine pixel-based features and by ANN and MDC with nine and 27 object-based features. The accuracy indices indicate that the ANN algorithm produced better results using object-based (OAA = 76.64%, kappa = 0.7071) than pixel-based classification (72.79%, 0.6687). When we compare the effectiveness of different numbers of feature variables on the object-based classification accuracy by ANN or MDC, mapping urban land cover using more input features (27) was better than that using fewer features (nine). The evaluation results in table 6 also show that ANN outperformed MDC using the same number of object-based input features. These evaluation results are basically consistent with those analysed with accuracy indices derived from test samples from the previous sections (table 4).

4.5 *Comparison*

Based on the accuracy indices derived from the test samples (averaged from the three sets of test samples) and the evaluation results derived from the 441 evaluation points, a comparison analysis was conducted from the following three aspects. First, to

Table 5. Confusion matrices created by ANN using IO test samples with (A) 27 features and (B) nine features.

	Reference										
	BT	NT	PT	Sh	GL	BR	IA	SS	Wa	Sum	UA (%)
(A) Confusion matrix created from the second set of test samples with 27 features.											
Classified											
BT	96	3	4	6	0	0	0	0	0	109	88.07
NT	6	20	1	0	1	0	0	0	0	28	71.43
PT	3	5	17	3	0	0	0	0	0	28	60.71
Sh	8	0	1	18	1	0	0	0	0	28	64.29
GL	0	0	0	2	28	0	0	0	0	30	93.33
BR	0	0	0	0	0	126	25	1	0	152	82.89
IA	0	0	0	0	0	31	115	2	2	150	76.67
SS	0	0	0	0	0	0	2	30	0	32	93.75
Wa	0	0	0	0	0	0	0	0	22	22	100.00
Sum	113	28	23	29	30	157	142	33	24	579	
PA (%)	84.96	71.43	73.91	62.07	93.33	80.25	80.99	90.91	91.67		
		AA = 81.06%		OAA = 81.52%		kappa = 0.7731		variance = 0.000415			
(B) Confusion matrix created from the second set of test samples with nine features.											
Classified											
BT	84	4	4	6	1	0	0	0	0	99	84.85
NT	9	18	3	5	1	0	0	0	0	36	50.00
PT	6	4	12	0	1	0	0	0	0	23	52.17
Sh	13	2	3	16	0	0	0	0	0	34	47.06
GL	1	0	1	2	27	0	0	0	0	31	87.10
BR	0	0	0	0	0	114	8	0	0	122	93.44
IA	0	0	0	0	0	40	125	0	0	165	75.76
SS	0	0	0	0	0	2	9	33	0	44	75.00
Wa	0	0	0	0	0	1	0	0	24	25	96.00
Sum	113	28	23	29	30	157	142	33	24	579	
PA (%)	74.34	64.29	52.17	55.17	90.00	72.61	88.03	100.00	100.00		
		AA = 77.40%		OAA= 78.24%		kappa = 0.7356		variance = 0.000450			

PA, producer's accuracy; UA, user's accuracy.

Table 6. Evaluation results obtained by comparing mapped results with high-resolution photo-interpreted results at 441 grid points. The mapped results are from two classification units (pixel- and object-based) and two algorithms (ANN and MDC) with nine or 27 features (bands).

	27fIO-ANN		9fIO-ANN		27fIO-MDC		9fIO-MDC		9fPIX-ANN	
	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
BT	89.31	95.95	89.94	97.95	86.16	90.13	77.99	89.21	62.26	97.06
NT	Inf.	0.00	Inf.	0.00	Inf.	0.00	Inf.	0.00	Inf.	0.00
PT	50.00	10.00	50.00	10.00	50.00	6.67	0.00	0.00	100.00	5.26
Sh	65.00	56.52	35.00	77.78	45.00	75.00	45.00	60.00	45.00	36.00
GL	71.74	94.29	73.91	91.89	65.22	100.00	54.35	92.59	73.91	82.93
BR	91.67	70.51	66.13	54.67	80.00	56.47	68.33	46.59	58.33	76.09
IA	76.00	92.68	60.20	77.63	64.00	81.01	51.00	71.83	92.00	79.31
SS	100.00	82.22	97.37	74.00	94.59	94.59	86.49	84.21	89.19	86.84
Wa	94.12	100.00	100.00	94.12	94.12	100.00	94.12	100.00	100.00	89.47
OAA (%)	84.58		76.64		77.10		67.57		72.79	
Kappa	0.8053		0.7071		0.7108		0.5977		0.6687	

PA, producer's accuracy; UA, user's accuracy; 27fIO or 9fIO, with 27 or nine selected IO-based features (bands); 9fPIX, with nine pixel-based features (bands); ANN, artificial neural network; MDC, minimum distance classifier.

consider the two types of classification units (pixel-based and object-based), we compared the accuracy indices produced from the test samples and evaluation points for the pixel-based classification with that from the object-based classification using the ANN algorithm with nine input features. The test sample results presented in table 4 show that all of the accuracy indices (AA, OAA and kappa) produced with the object-based data were consistently higher than those with the pixel-based image data. The evaluation results presented in table 6 (nine selected IO-based (9&IO)-ANN vs. nine pixel-based (9&PIX)-ANN) indicate that the resultant map created with the object-based data is better than that with the pixel-based data. Second, in considering the performance of the two algorithms (ANN vs. MDC) with object-based features (table 4), ANN significantly outperformed the MDC algorithm when using the same object-based features as inputs. The OAA and kappa values listed in table 6 also clearly show that ANN outperformed MDC. Third, comparing the effects of different numbers of features on the urban land cover classification reveals that all of the accuracy indices (AA, OAA and kappa) with 27 object-based features were higher than those with nine object-based features (tables 4–6).

To test whether these differences (between the two classification units, the two algorithms, and the two numbers of features) were statistically significant, Z-statistics were calculated from the Kappa value and corresponding variance derived from the test samples and are presented in table 7 (Z-statistics were not calculated from the 441 evaluation results because evaluation points are not available or are not reliable for land cover classes NT and PT). From table 7, Z-statistics were calculated to compare the two classification units using the ANN algorithm (9fIO-ANN vs. 9fPIX-ANN), between the two numbers of selected features (27fIO-ANN vs. 9fIO-ANN and 27fIO-MDC vs. 9fIO-MDC), and between the ANN and MDC with the same numbers of object-based features (27fIO-ANN vs. 27fIO-MDC and 9fIO-ANN vs. 9fIO-MDC).

Table 7. Z-statistic tests calculated from kappa-variance of classification results of test samples, generated with different classification units (pixel- and object-based) and different algorithms (ANN and MDC) with nine or 27 features.

	27fIO-ANN	9fIO-ANN	27fIO-MDC	9fIO-MDC	9fPIX-ANN
27fIO-ANN	n/a				
9fIO-ANN	1.242	n/a			
27fIO-MDC	1.478	n/a	n/a		
9fIO-MDC	n/a	2.819***	2.556**	n/a	
9fPIX-ANN	n/a	1.886*	n/a	n/a	n/a

*Difference between classification accuracies is significant at the 0.90 confidence level.
**Difference between classification accuracies is significant at the 0.95 confidence level.
***Difference between classification accuracies is significant at the 0.99 confidence level.
$Z = |k_1 - k_2| / \sqrt{v_1 + v_2}$ where k_1 and k_2 are the kappa values of the input features 1 and 2, respectively, and v_1 and v_2 are the corresponding variances.
For different classification units and names of the algorithms, see table 6.

The Z-statistics indicate a significant difference ($p < 0.1$) of ANN results produced with different classification units (pixel-based and object-based). Although the absolute accuracy indices derived using ANN with 27 input features are all higher than those with nine input features, the difference of results created using ANN with different numbers of features (9 vs. 27) was not significant even at the 0.90 confidence level. However, the difference of results created using the MDC algorithm with different numbers of features (9 vs. 27) was significant ($p < 0.05$). The difference of results between ANN and MDC with the nine features was statistically significant at the 0.99 confidence level, but the difference between ANN and MDC with 27 features was not significant even at the 0.90 confidence level. Despite the improved performance using the 27 input features with MDC, the results related to ANN with different numbers of features imply that there is still a lot of redundant information among the 27 feature variables that did not proportionally improve the classification results compared to using the nine input features.

5. Discussion

Our experimental results have demonstrated that the object-based urban land cover classification outperformed the pixel-based classification. When we used the same input feature variables (i.e. nine features) and classifier (i.e. ANN) but different classification units (i.e. pixel-based and object-based data), the improvement in the classification result with the object-based classification unit was statistically significant at the 0.90 confidence level compared to the pixel-based unit (AA increasing by 3.1%, OAA increasing by 4.7% and kappa increasing by 0.04; from the evaluation results, OAA increasing by 3.8% and kappa increasing by 0.04). Unlike the pixel-based techniques, which use pixel values and kernel-based texture/contexture information, the object-based techniques can also use geometric/shape and contextual information calculated from individual IOs of a scene covering the study area. Therefore, by including more features (27) and some textural and shape/geometric features only available to IOs, we achieved an improvement in the classification results compared to the pixel-based technique and fewer input features (AA increasing by 7.2%, OAA increasing by 7.7% and kappa increasing by 0.08; from the evaluation results, OAA increasing by 11.8% and kappa increasing by 0.14). Some features derived from IOs were efficient in distinguishing grass/lawn from tree canopies (figure 4) and separating BR and IA surface areas (figure 5). The textural feature shown in figure 4 was effective in separating the relatively coarse tree canopies from the relatively smooth grass/lawn. Among the five features shown in the table in figure 5, three features are related to 'spectral variation' and two are related to 'colour'

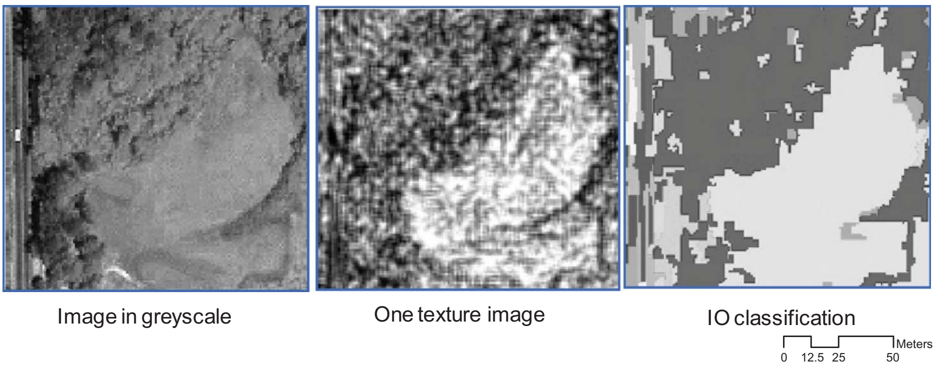
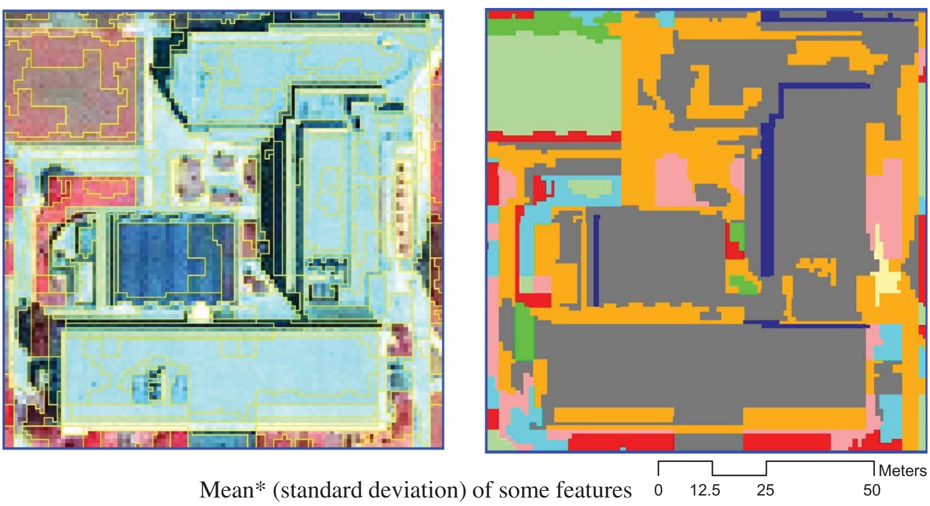


Figure 4. Illustrating the utility of image texture information to separate tree canopy from grass/lawn area.



Mean* (standard deviation) of some features

Features	Sat	SDB4	SDH	SDVI	GLCMH
Roof	1844.07 (777.59)	244.21 (111.33)	482.14 (222.61)	482.14 (25.11)	0.057 (0.031)
Ground Impervious	3000.00 (675.81)	359.44 (136.23)	297.59 (148.77)	237.19 (136.62)	0.030 (0.012)

*All feature mean differences between two components are significant at $\alpha = 0.05$, except $p = 0.053$ for SDB4.

Figure 5. Some features help to identify ground impervious surface from building roof. The embedded table shows means and standard deviations of five features: Sat, SDB4, SDH, SDVI and GLCMH (definitions given in table 1). Ten samples (IOs) were used for calculating the means and corresponding standard deviations.

saturation and texture information, which may be important when identifying BR and IA classes. Considering the integrative effectiveness of spectral, textural and geometric/shape information extracted from IOs for classifying urban detailed land cover, our results suggest that classification with object-based data should have advantages over that of pixel-based data. Such a conclusion is not surprising if we also consider previous studies conducted by other researchers. For example, Shackelford and Davis (2003), Yu *et al.* (2006) and Guo *et al.* (2007) used object-based high spatial resolution imagery (airborne or satellite image data) to obtain similar conclusions, including: improved identification of urban surface components; increased accuracy of vegetation community classification; and more accurately mapped oak tree mortality. In addition, when some researchers compared object-based techniques with pixel-based techniques for change detection analysis (i.e. deforestation analysis and other land cover change analysis), analysis accuracies were improved significantly (Al-Khudhairy *et al.* 2005, Desclée *et al.* 2006, Zhou *et al.* 2008). Our experimental results as well as these previous studies demonstrate the efficiency of using object-based classification techniques over that of using pixel-based approaches.

Our object-based classification results indicate that ANN outperformed MDC using the same number of input features. We suggest that the reason for this improved performance was that ANN made use of subtle spectral differences in the four PS band images more efficiently because of the ANN's multilayer structure and its non-parametric properties. Previous research has also reported that ANN produced better classification results with remote sensing data when compared with traditional methods, such as the MDC (e.g. Gong *et al.* 1997, Erbek *et al.* 2004).

Using the ANN algorithm, the difference in results created with different numbers of input features (9 vs. 27) was not significant at the 0.90 confidence level even though all accuracy indices derived from the 27-feature input were higher than those from the nine-feature input. This may be explained in two ways: first, the large (accuracy index) variances with both nine and 27 features, relative to those with pixel-based (e.g. 0.000030 vs. 0.000454 for the nine features), tend to decrease the significant level of difference of the classification accuracy. Such high variances might have resulted because some class accuracies were much higher than others, as shown in table 5 (e.g. 100% producer's accuracies of SS and Wa vs. less than 55% of PT and Sh with the nine-feature, and more than 92% producer's accuracies of GL and Wa vs. lower than 71% of NT and Sh with the 27-feature). Second, although we used an SDA to select a subset from all candidate feature variables, there was likely to be redundant information remaining among the selected 27 feature variables such that some features made a relatively low contribution to separability among classes (e.g. those features with an *F*-value <10 in table 3), and did not proportionally improve the classification results with an increased number of feature variables. In this study, the use of SAVI and three ratio features (i.e. Ratio1 to Ratio3 selected into the subset of 27 features), based on their definitions, should have weakened the effect of the shadow on the classification results. However, the effect of the shadow on the mapped results was evident. Considering both the SDA method and MDC use a linear discriminant property, unlike ANN, which is non-linear, it should be reasonable that the 27 selected features resulted in a significant accuracy improvement compared to the results created with nine features by MDC.

6. Summary and conclusions

In this study, we tested whether the IO technique could significantly improve classification accuracy compared to the pixel-based method when applied to urban detailed land cover mapping with high spatial resolution IKONOS imagery in Tampa Bay, FL, USA. For this case study, we first carried out IKONOS image data calibration and data fusion with the PS process, then extracted and used nine features (four PS bands, three HIS transferred indices, one SAVI and one textural image) as basic input data layers to conduct pixel-based urban land cover classification directly and to perform image segmentation to create IOs for an object-based classification. An SDA was used to select 27 important IO features, and then two algorithms (ANN and MDC) were used to map urban land cover types with either pixel- or object-based data. Finally, the urban land cover mapping results were evaluated with visually interpreted results from high-resolution (0.3-m) digital aerial photographs.

Our results indicate a statistically significant difference in accuracy between the pixel- and object-based classification of urban detailed land surface components using high spatial resolution IKONOS data. This is because object-based input features eliminate the 'salt-and-pepper' effect on classification through image segmentation to create IOs, using spectral, textural/contextual and shape/geometric features extracted from the IOs. We evaluated the performance of the two algorithms with different numbers of object-based features and showed that ANN outperformed MDC in the urban land cover classification. This is possible because ANN can handle non-perfect parametric image features such as SAVI and textural features and efficiently utilize some texture and shape/geometric features extracted from the IOs. According to the experimental results, a non-parametric and non-linear classifier such as ANN should be used extensively when conducting an object-based classification with inputs of spectral, textural/contextual and geometric features extracted from IOs. This study also proved that using more input features (27 vs. 9 features) could improve the IO classification accuracy by both ANN and MDC, and especially by MDC, but was not statistically significant at the 0.90 confidence level for ANN. This might be attributed to redundant information still existing among the selected features and possibly to the impact of shadow. Some issues related to image segmentation worthy of greater attention for future research include: how to select the appropriate criteria to create ideal IOs to achieve accuracy for a particular application; how to evaluate whether edge and shape of IOs overlap (coincide with) the landscape boundaries (land cover type/patch) through justifying scales; and operationally, what relationship exists between IOs and ecological units. These issues should be considered in developing object-based techniques with high spatial resolution imagery in the future.

Acknowledgements

This work was partially supported by the University of South Florida (USF) under the New Researcher Grant (no. 18300). We thank USF graduate student J. Kunzer for his help in the field trip, and Dr M. Andreu and M. Friedman for providing ground plot measurements from the City of Tampa Ecological Analysis 2006–2007.

References

AL-KHUDHAIRY, D.H.A., CARAVAGGI, I. and GIADA, S., 2005, Structural damage assessments from IKONOS data using change detection, object-oriented segmentation, and classification techniques. *Photogrammetric Engineering and Remote Sensing*, **71**, pp. 825–837.

ANDREU, M.G., FRIEDMAN, M.H., LANDRY, S.M. and NORTHROP, N.J., 2008, *City of Tampa Urban Ecological Analysis 2006–2007*. Final Report to the City of Tampa, 24 April 2008, City of Tampa, Florida. Available online at: http://www.sfrc.ufl.edu/urbanforestry/Files/TampaUEA2006-7_FinalReport.pdf (accessed 22 November 2010).

BAATZ, M., BENZ, U., DEHGHANI, S., HEYNEN, M., HÖLTJE, A., HOFMANN, P., LINGENFELDER, I., MIMLER, M., SOHLBACH, M., WEBER, M. and WILLHAUCK, G., 2004, *eCognition Professional User's Guide* (München, Germany: Definiens Imaging GmbH).

BAATZ, M. and SCHAPE, A., 2000, Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In *Angewandte Geographische Informations – Verarbeitung XII*, J. Strobl, T. Blaschke and G. Griesebner (Eds.), pp. 12–23 (Karlsruhe: Wichmann Verlag). Available online at: http://www.ecognition.cc/download/baatz_schaepe.pdf (accessed 22 November 2010).

BENZ, U.C., HOFMANN, P., WILLHAUCK, G., LINGENFELDER, I. and HYMEN, M., 2004, Multi-resolution, object oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, **58**, pp. 239–258.

BLASCHKE, T., 2010, Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, **65**, pp. 2–16.

CARLEER, A.P. and WOLFF, E., 2006, Region-based classification potential for land-cover classification with very high spatial resolution satellite data. In *Proceedings of the 1st International Conference on Object-Based Image Analysis (OBIA 2006)*, 4–5 July 2006, Salzburg University, Austria, Vol. 36, ISSN 1682–1777. Available online at: http://www.isprs.org/proceedings/XXXVI/4-C42/papers.htm (accessed 22 November 2010).

CLARK, M.L., ROBERTS, D.A. and CLARK, D.B., 2005, Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, **96**, pp. 375–398.

CONGALTON, R.G. and GREEN, K., 1999, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices* (New York: Lewis Publishers).

CONGALTON, R.G. and MEAD, R.A., 1983, A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing*, **49**, pp. 69–74.

DAVIS, C.H. and WANG, X., 2003, Planimetric accuracy of Ikonos 1m panchromatic orthoimage products and their utility for local government GIS basemap applications. *International Journal of Remote Sensing*, **24**, pp. 4267–4288.

DESCLÉE, B., BOGAERT, P. and DEFOURNY, P., 2006, Forest change detection by statistical object-based method. *Remote Sensing of Environment*, **102**, pp. 1–11.

ERBEK, F.S., ÖZKAN, C. and TABERNER, M., 2004, Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. *International Journal of Remote Sensing*, **25**, pp. 1733–1748.

GONG, P. and HOWARTH, P.J., 1990, The use of structural information for improving spatial resolution and classification accuracy land-cover classification accuracies at the rural–urban fringe. *Photogrammetric Engineering and Remote Sensing*, **56**, pp. 67–73.

GONG, P., PU, R. and YU, B., 1997, Conifer species recognition: an exploratory analysis of in situ hyperspectral data. *Remote Sensing of Environment*, **62**, pp. 189–200.

GUO, Q., KELLY, M., GONG, P. and LIU, D., 2007, An object-based classification approach in mapping tree mortality using high spatial resolution imagery. *GIScience and Remote Sensing*, **44**, pp. 24–47.

HARALICK, R.M., SHANMUGAM, K. and DINSTEIN, I., 1973, Textural features for image classification. *IEEE Transactions on Geoscience and Remote Sensing*, **3**, pp. 610–621.

HAY, G.J. and CASTILLA, G., 2008, Geographic object-based image analysis (GEOBIA): a new name for a new discipline. In *Object-Based Image Analysis*, T. Blaschke, S. Lang and G. Hay (Eds.), pp. 75–90 (New York: Springer).

HEROLD, M., LIU, X.H. and CLARKE, K.C., 2003, Spatial metrics and image texture for mapping urban land-use. *Photogrammetric Engineering and Remote Sensing*, **69**, pp. 991–1001.

HILL, R.A., 1999, Image segmentation for humid tropical forest classification in Landsat TM data. *International Journal of Remote Sensing*, **20**, pp. 1039–1044.

ITT, 2006, *ENVI Version 4.3* (Boulder, CO: ITT Industries, Inc.).

JAIN, S. and JAIN, R.K., 2006, A remote sensing approach to establish relationships among different land covers at the micro level. *International Journal of Remote Sensing*, **27**, pp. 2667–2682.

JENSEN, J.R., 2005, *Introductory Digital Image Processing: A Remote Sensing Perspective*, 3rd edn (Upper Saddle River, NJ: Prentice Hall), pp. 210–215.

JOHNSSON, K., 1994, Segment-based land-use classification from SPOT satellite data. *Photogrammetric Engineering and Remote Sensing*, **60**, pp. 47–53.

KONG, C., KAI, X. and WU, C., 2006, Classification and extraction of urban land-use information from high-resolution image based on object multi-features. *Journal of China University of Geosciences*, **17**, pp. 151–157.

KOSAKA, N., AKIYAMA, T., BIEN, T. and KOJIMA, T., 2005, Forest type classification using data fusion of multispectral and panchromatic high-resolution satellite imageries. *Geoscience and Remote Sensing Symposium Proceedings, 2005. IGARSS '05. IEEE International*, **4**, pp. 2980–2983.

KUX, H.J. and ARAÚJO, E.H.G., 2008, Object-based image analysis using QuickBird satellite images and GIS data, case study Belo Horizonte (Brazil). In *Object-Based Image Analysis*, T. Blaschke, S. Lang and G. Hay (Eds.), pp. 571–588 (New York: Springer).

LALIBERTE, A.S., RANGO, A., HAVSTAD, K.M., PARIS, J.F., BECK, R.F., MCNEELY, R. and GONZALEZ, A.L., 2004, Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico. *Remote Sensing of Environment*, **93**, pp. 198–210.

LATTY, R.S., NELSON, R., MARKHAM, B., WILLIAMS, D., TOLL, D. and IRONS, J., 1985, Performance comparison between information extraction techniques using variable spatial resolution data. *Photogrammetric Engineering and Remote Sensing*, **51**, pp. 1159–1170.

MARCHESI, A., COLOMBO, R. and VALENTINI, P., 2006, Application of high spatial resolution satellite imagery for urban environment mapping. In *Proceedings of the 1st International Conference on Object-Based Image Analysis (OBIA 2006)*, 4–5 July 2006, Salzburg University, Austria, Vol. 36, ISSN 1682–1777. Available online at: http://www.isprs.org/proceedings/XXXVI/4-C42/papers.htm (accessed 22 November 2010).

MARTIN, L.R.G., HOWARTH, P.J. and HOLDER, G., 1988, Multispectral classification of land use at the rural–urban fringe using SPOT data. *Canadian Journal of Remote Sensing*, **14**, pp. 72–79.

MATHIEU, R., ARYAL, J. and CHONG, A.K., 2007, Object-based classification of IKONOS imagery for mapping large-scale vegetation communities in urban areas. *Sensors*, **7**, pp. 2860–2880.

MUCHONEY, D.M. and STRAHLER, A.H., 2002, Pixel- and site-based calibration and validation methods for evaluating supervised classification of remotely sensed data. *Remote Sensing of Environment*, **81**, pp. 290–299.

NICHOL, J. and WONG, M.S., 2007, Remote sensing of urban vegetation life form by spectral mixture analysis of high-resolution IKONOS satellite images. *International Journal of Remote Sensing*, **28**, pp. 985–1000.

PAO, Y., 1989, *Adaptive Pattern Recognition and Neural Networks* (New York: Addison-Wesley).

Pu, R., Kelly, M., Anderson, G.L. and Gong, P., 2008, Using CASI hyperspectral imagery to detect mortality and vegetation stress associated with a new hardwood forest disease. *Photogrammetric Engineering and Remote Sensing*, **74**, pp. 65–75.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1986, Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, pp. 318–362 (Cambridge, MA: MIT Press).

Santos, T., Tenedório, J.A., Encarnação, S. and Rocha, J., 2006, Comparing pixel vs. object based classifiers for land cover mapping with Envisat-MERIS data. In *26th EARSeL Symposium*, Maio, Varsóvia, pp. 1–9. Available online at: http://egeo.fcsh.unl.pt/pdf/linhamtig_santos_tenedorio_encarnacao_rocha_earsel_varsovia2006.pdf. (accessed 22 November 2010).

SAS, 1991, *SAS/STA User's Guide*, *Release 6.03* (Cary, NC: SAS Institute Inc.).

Schowengerdt, R.A., 2007, *Remote Sensing: Models and Methods for Image Processing*, 3rd edn, pp. 387–455 (New York: Academic Press).

Shackelford, A.K. and Davis, C.H., 2003, A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, pp. 2354–2363.

Song, C., 2005, Spectral mixture analysis for subpixel vegetation fractions in the urban environment: how to incorporate endmember variability? *Remote Sensing of Environment*, **95**, pp. 248–263.

Story, M. and Congalton, R., 1986, Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, **52**, pp. 397–399.

SWFWMD, 2006, *2006 Natural Color Imagery for Hillsborough County*. Southwest Florida Water Management District, Brooksville, FL. Available online at: www.swfwmd.state.fl.us/data/gis/doqq_search.htm (accessed 2 January 2007).

Ton, J.C., Sticklen, J. and Jain, A.K., 1991, Knowledge-based segmentation of Landsat images. *IEEE Transactions on Geoscience and Remote Sensing*, **29**, pp. 222–232.

Townshend, J. and Justice, J., 1981, Information extraction from remotely sensed data, a user view. *International Journal of Remote Sensing*, **2**, pp. 313–329.

Townsend, P.A. and Walsh, S.J., 2001, Remote sensing of forested wetlands: application of multitemporal and multispectral satellite imagery to determine plant community composition and structure in southeastern USA. *Plant Ecology*, **157**, pp. 129–149.

Treitz, P. and Howarth, P.J., 2000, High spatial resolution remote sensing data for forest ecosystem classification: an examination of spatial scale. *Remote Sensing of Environment*, **72**, pp. 268–289.

van Aardt, J.A.N. and Wynne, R.H., 2007, Examining pine spectral separability using hyperspectal data from an airborne sensor: an extension of field-based results. *International Journal of Remote Sensing*, **28**, pp. 431–436.

Woodcock, C.E. and Strahler, A.H., 1987, The factor of scale in remote sensing. *Remote Sensing of Environment*, **21**, pp. 311–332.

Yang, L., Xian, G., Klaver, J.M. and Deal, B., 2003, Urban land-cover change detection through sub-pixel imperviousness mapping using remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, **69**, pp. 1003–1010.

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

Zhou, W., Troy, A. and Grove, M., 2008, Object-based land cover classification and change analysis in the Baltimore metropolitan area using multitemporal high resolution remote sensing data. *Sensors*, **8**, pp. 1613–1636.