International Journal of Remote Sensing
Vol. 34, No. 7, 10 April 2013, 2583–2606

Taylor & Francis
Taylor & Francis Group

GeoEye-1 and WorldView-2 pan-sharpened imagery for object-based classification in urban environments

M.A. Aguilar*, M.M. Saldaña, and F.J. Aguilar

Departamento de Ingeniería Rural, University of Almería, Escuela Superior de Ingeniería, 04120 Almería, Spain

(*Received 11 June 2012; accepted 5 October 2012*)

The latest breed of very high resolution (VHR) commercial satellites opens new possibilities for cartographic and remote-sensing applications. In fact, one of the most common applications of remote-sensing images is the extraction of land-cover information for digital image base maps by means of classification techniques. The aim of the study was to compare the potential classification accuracy provided by pan-sharpened orthoimages from both GeoEye-1 and WorldView-2 (WV2) VHR satellites over urban environments. The influence on the supervised classification accuracy was evaluated by means of an object-based statistical analysis regarding three main factors: (i) sensor used; (ii) sets of image object (IO) features used for classification considering spectral, geometry, texture, and elevation features; and (iii) size of training samples to feed the classifier (nearest neighbour (NN)). The new spectral bands of WV2 (Coastal, Yellow, Red Edge, and Near Infrared-2) did not improve the benchmark established from GeoEye-1. The best overall accuracy for GeoEye-1 (close to 89%) was attained by using together spectral and elevation features, whereas the highest overall accuracy for WV2 (83%) was achieved by adding textural features to the previous ones. In the case of buildings classification, the normalized digital surface model computed from light detection and ranging data was the most valuable feature, achieving producer's and user's accuracies close to 95% and 91% for GeoEye-1 and VW2, respectively. Last but not least and regarding the size of the training samples, the rule of 'the larger the better' was true but, based on statistical analysis, the ideal choice would be variable depending on both each satellite and target class. In short, 20 training IOs per class would be enough if the NN classifier was applied on pan-sharpened orthoimages from both GeoEye-1 and WV2.

1. Introduction

The successful launch of the first very high resolution (VHR) satellites, such as IKONOS in September 1999 or QuickBird in October 2001, marked the beginning of a wholly new age in remote sensing. In fact, VHR satellites are able to capture images of the Earth's surface with a ground sample distance (GSD) of 1 m and even less. Currently, GeoEye-1, launched in 2008, is the commercial VHR satellite with the highest geometric resolution in both panchromatic (PAN) and multispectral (MS) products (nominal GSD at nadir of 0.41 and 1.65 m, respectively). More recently, on 4 January 2010, VHR satellite imagery of WorldView-2 (WV2) with 0.46 and 1.84 m nominal resolution at nadir in PAN and MS, respectively, started to become available commercially. However, image products from

*Corresponding author. Email: maguilar@ual.es

ISSN 0143-1161 print/ISSN 1366-5901 online

http://dx.doi.org/10.1080/01431161.2012.747018
http://www.tandfonline.com

GeoEye-1 and WV2 have to be down-sampled to 0.5 and 2 m in PAN and MS, respectively, for commercial sales, as a requirement levied by the US Government. Nevertheless, and without any doubt, the most relevant technical innovation of WV2 lies in an improvement of its potential spectral performance since the number of bands that comprise its MS product has been increased to 8, instead of the four classic bands (Blue, Green, Red, and Near Infrared) offered by all previous VHR satellites. The new spectral bands of WV2 (Coastal, Yellow, Red Edge, and Near Infrared-2) could provide a significant improvement in classification accuracy ranging from 5% to 20% for certain land-cover types such as man-made materials, selected vegetation targets, soils, shadows, turbidity in water, and deeper bathymetry (Marchisio, Pacifici, and Padwick 2010). On the other hand, and working with WV2 imagery for land-cover identification of hazelnuts fields, Taşdemir and Reis (2011) reported that the overall accuracy was surprisingly higher considering only the four traditional bands instead of the available eight bands. In this regard, Marshall, Lewis, and Ostendorf (2012) concluded that there was no benefit in using the additional four bands of WV2 to discriminate invasive grass species.

Furthermore, the evolution and dynamism of new man-made infrastructure, houses, and buildings requires a frequent updating of urban cartographic databases. Traditionally, this specific task has been carried out by photointerpretation of aerial images, which is expensive and time consuming (e.g. Hermosilla et al. 2011). However, many recent studies have used different VHR satellite imagery for automatically extracting georeferenced data in urban environments such as IKONOS (e.g. Turker and San 2010; Pu, Landry, and Yu 2011), GeoEye-1 (e.g. Grigillo and Fras 2011; Hussain et al. 2011), and WV2 (e.g. Longbotham et al. 2012; Zhou et al. 2012). In fact, automatic building extraction or classification from VHR is a very challenging task and has been the focus of intensive research for the last decade. In this way, a higher local variance of urban land-cover classes is found when the resolution of the input image is increased (Myint et al. 2011), and therefore the accuracy of the traditional pixel-based classification approaches is reduced and the results can show a 'salt and pepper' effect (Treitz and Howarth 2000; Pu, Landry, and Yu 2011). 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, Aryal, and Chong 2007). To overcome this problem, object-based image analysis (OBIA) has been pointed out as a very interesting option to improve the performance of supervised classifiers (Carleer and Wolff 2006; Im, Jensen, and Tullis 2008; Blaschke 2010; Lu, Hetrick, and Moran 2010; Myint et al. 2011). In fact, it is a new paradigm in the field of geographic information science, in which images are segmented into meaningful segments (or objects) according to different criteria before classification is carried out. OBIA techniques are based on aggregating similar pixels to obtain homogenous objects, which are then assigned to a target class. Using objects instead of pixels as a minimum unit of information minimizes the salt and pepper effect due to the spectral heterogeneity of individual pixels. Furthermore, and unlike traditional pixel-based methods that only use spectral information, object-based approaches can use shape, texture, and context information associated with the objects and thus have the potential to efficiently handle more difficult image analysis tasks (e.g. Marpu et al. 2010). Thus, the number of available OBIA works is increasing rapidly, and they are focusing on subtopics such as specific OBIA hierarchy and scale concepts (Addink, de Jong, and Pebesma 2007), image segmentation (Drăguţ, Tiede, and Levick 2010; Marpu et al. 2010), OBIA change detection (Im, Jensen, and Tullis 2008), and OBIA accuracy assessment (Platt and Rapoza 2008; Albrecht, Lang, and Hölbling 2010; Lang et al. 2010). A comprehensive review of the advantages and disadvantages of using OBIA techniques for image classification, as well as the state of the art of these methods, can be found in Blaschke (2010).

Bearing in mind that supervised classification is one of the most commonly undertaken analyses in remote sensing, it is important to note that the attained classification accuracy is a function of the training data used, mainly size and quality (e.g. Foody 1999; Tsai and Philpot 2002; Foody and Mathur 2006). In this way, classifiers can need different kinds of training information (Foody 1999), i.e. a particular training set can produce a highly accurate classification from one classifier and may yield a considerably lower accuracy if used with another classifier. The literature, however, generally pleads for a relatively uniform or classifier-independent approach regarding training set design. Guidance on the design of the classification training phase typically calls for the use of large samples (the larger the better) of randomly selected pure or meaningful objects to characterize the corresponding target classes. It is common practice in remote sensing to ensure that the number of training samples for each class should comprise at least 10–30 times the number of wavebands used in the analysis (Piper 1992; Mather 2004). However, the nature of an ideal training set is unclear and it mainly depends on the aim of the training stage.

In this article, object-based supervised classification accuracies in urban environments have been tested from both GeoEye-1 and WV2 pan-sharpened orthoimages captured almost in the same conditions. The variables studied were the following: (i) sensor used, (ii) sets of image object features used for classification (features vector), and (iii) percentage of training samples considered. A statistical test was carried out to strengthen the conclusions.

2. Study site and data sets

2.1. Study site

The study area comprises 17 ha in the little seaside village of Villaricos, located in the province of Almería, southern Spain (Figure 1). Its urban landscape presents high heterogeneity, mixing old buildings and new housing developments.

2.2. GeoEye-1 orthoimages

One image of GeoEye-1 Geo acquired over the study site was captured in reverse scan mode on 29 September 2010, simultaneously recording the PAN band (spectral range from 450 to 800 nm) and all four MS bands: Blue (450–510 nm), Green (510–580 nm), Red

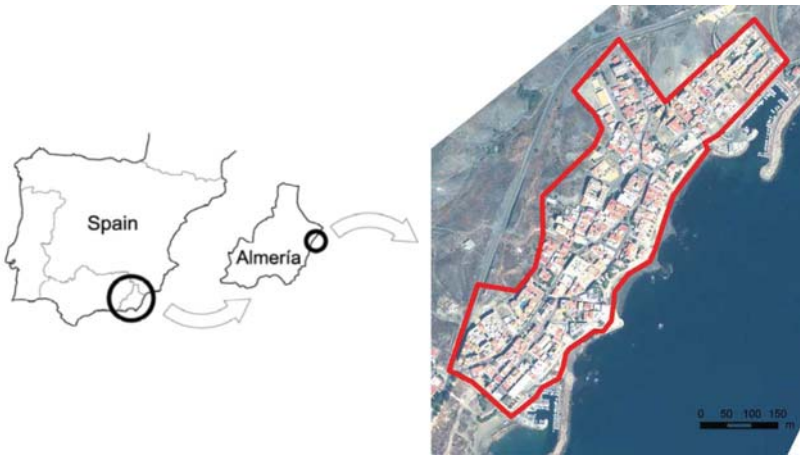

Figure 1. Location of the working area. Coordinates WGS84 37.2489° N, 1.7709° W.

(655–690 nm), and Near Infrared (780–920 nm). It was ordered with a dynamic range of 11 bits per pixel and without any adjustment (i.e. maintaining absolute radiometric accuracy and full dynamic range for scientific applications). Moreover, and due to US Government Licensing, it was resampled to 0.5 and 2 m for the PAN and MS cases, respectively. The corresponding pan-sharpened image, with 0.5 m GSD and containing the spectral information gathered from the MS image (four bands), was attained by using the PANSHARP module included in Geomatica v. 12 (PCI Geomatics, Richmond Hill, ON, Canada). Lastly, two 0.5 m GSD orthoimages (PAN and pan-sharpened) were computed by using OrthoEngine, the photogrammetric module of Geomatica. Both orthoimages were obtained by applying the rational function model with zero-order transformation in image space and using very accurate ancillary data such as seven ground control points (GCPs) measured by DGPS and a lidar (light detection and ranging) derived digital elevation model (DEM), which will be explained in depth in Section 2.4. These orthoimages presented a two-dimensional root mean square error ($RMSE_{2d}$) of 0.46 m computed on 75 independent check points (Aguilar et al. 2012a). Other characteristics of the employed GeoEye-1 Geo bundle are shown in Table 1.

2.3. WV2 orthoimages

One Ortho Ready Standard Level-2A (ORS2A) bundle image (PAN + MS) of WV2 was acquired on 18 August 2011, covering the entire study area (Table 1). Its corresponding 0.5 m PAN image presented a spectral range from 450 to 800 nm, whereas its 2 m MS image was composed of eight bands: Coastal Blue (400–450 nm), Blue (450–510 nm), Green (510–580 nm), Yellow (585–625 nm), Red (630–690 nm), Red Edge (705–745 nm), Near Infrared-1 (760–895 nm), and Near Infrared-2 (860–1040 nm). The delivered products presented a dynamic range of 11 bits. Any colour correction or contrast enhancement was carried out. Finally, PAN and pan-sharpened (eight bands) orthoimages with 0.5 m GSD were generated from WV2 following the same aforementioned procedure that was applied to GeoEye-1. The $RMSE_{2d}$ planimetric accuracy attained for both products was 0.46 m (Aguilar, Saldaña, and Aguilar 2012b).

Table 1. Characteristics of the bundle images from GeoEye-1 Geo and WV2 ORS2A acquired at the study site.

Product	GeoEye-1 Geo PAN	GeoEye-1 Geo 4 band MS	WV2 ORS2A PAN	WV2 ORS2A 8 band MS
Acquisition date	29 September 2010		18 August 2011	
Acquisition time (GTM)	11:01		11:23	
Cloud cover	0 %		0%	
Scan direction	Reverse		Reverse	
Sun azimuth	159.3°		152.8°	
Sun elevation	48.4°		63.8°	
Collection elevation	69.4°		80.0°	
Collection azimuth	221.9°		216.1°	
Collected col GSD	0.46 m	1.84 m	0.47 m	1.89 m
Collected row GSD	0.45 m	1.80 m	0.48 m	1.90 m
Product pixel size	0.5 m	2 m	0.5 m	2 m

2.4. Normalized digital surface model

The highly accurate elevation data used in this work, presenting 1 m grid spacing, were generated from lidar data taken on 28 August 2009 through a combined photogrammetric and lidar survey at a flying height above ground of approximately 1000 m. The Leica ALS60 airborne laser scanner (35 degrees field of view, FOV) was used with the support of a nearby ground GPS reference station to capture on average around 1.61 points m^{-2}. The estimated vertical accuracy computed on 62 open-terrain independent check points took a value of 8.9 cm. The corresponding normalized digital surface model (nDSM) was generated by subtracting DEM from DSM. In this way, buildings can be easily distinguished (Figure 2(*c*)). At the same time, the digital mapping camera Intergraph Z/I Imaging DMC was used for collecting RGB + Nir VHR digital images from which 0.15 m GSD orthoimages were attained.

3. Methodology

3.1. Image segmentation

Image segmentation is a crucial step of OBIA that splits an image into separated and homogeneous regions (objects). There exist several types of image segmentation algorithms developed for a variety of applications in various fields of image analysis. Most of them largely depend on specified parameters, which implies that segmentation is not an easy task. At this point, it should be clearly stated that much of the work referred to as OBIA has been originated around the software eCognition. Indeed, about 50–55% of the papers related to OBIA used this software package (Blaschke 2010). Thus, the OBIA software used in this research was eCognition v. 8.0. It uses a segmentation algorithm known as multiresolution segmentation, which is based on the Fractal Net Evolution Approach (Baatz and Schäpe 2000), and has achieved good segmentation accuracy (Marpu et al. 2010). This segmentation approach consists of a bottom-up region-merging technique starting with one-pixel objects. In numerous iterative steps, smaller image objects (IOs) are merged into larger ones (Baatz and Schäpe 2000). The multiresolution segmentation algorithm is an optimization procedure that, for a given number of IOs, minimizes the average heterogeneity and maximizes their respective homogeneity. The outcome of this segmentation algorithm is controlled by three main factors: (i) the homogeneity criteria or scale parameter that

(*a*) (*b*) (*c*)

Figure 2. Detailed information (110 m × 135 m) attained on the study area: (*a*) pan-sharpened orthoimage from GeoEye-1, (*b*) pan-sharpened orthoimage from WV2, and (*c*) nDSM.

determines the maximum allowed heterogeneity for the resulting IOs, (ii) the weight of colour and shape criteria in the segmentation process, and (iii) the weight of the compactness and smoothness criteria (i.e. the higher the compactness weight, the more compact IOs may be). The optimal determination of these three somewhat abstract terms is not easy to carry out. Traditionally, a systematic trial-and-error approach validated by the visual inspection of the quality of the output IOs would be required for setting these parameters (Mathieu, Aryal, and Chong 2007; Tian and Chen 2007). Moreover, the configuration of the optimal parameter depends on the desired objects to be segmented (e.g. Tian and Chen 2007). Recently there have emerged new tools for a fast estimation of the scale parameters of multiresolution segmentation (Drăguţ, Tiede, and Levick 2010), as well as for evaluating the final segmentation accuracy (Marpu et al. 2010).

However, and although accepting that VHR satellite imagery segmentation is a fundamental step for attaining high accuracies at the final classification (Song, Civco, and Hurd 2005; Liu and Xia 2010), we have to clarify that the goal of this article was strictly focused on comparing classification accuracy between pan-sharpened orthoimages taken from GeoEye-1 and WV2 satellites. Therefore, the segmentation step only had to assure pure objects (i.e. grouping pixels belonging to an only class) to later classify them and assess the final classification accuracy by using a ground truth based on IOs. Note that it was necessary to count on two manually delineated ground truths, one for each satellite, due to changes mainly on the shadow's position (see Section 3.2). Despite this, an attempt was made to achieve a similar number of IOs both for GeoEye-1 and WV2 orthoimages. In that sense, and by means of a trial-and-error process, 2723 pure IOs (Figure 3(*a*)) were finally segmented from GeoEye-1's orthoimages, which, after visual inspection, matched well with the feature boundaries of the land-cover types in the study area (mainly buildings). In the same way, 2720 pure IOs were finally extracted in the case of WV2's test (Figure 3(*b*)). In both tests, the multiresolution segmentation approach was used in two iterative steps. Furthermore, the compactness criterion was assigned a weight of 0.5 and the shape value was fixed at 0.3 (weight of colour = 0.7) in all the segmentations carried out.

For the pan-sharpened orthoimage from GeoEye-1, the first step was undertaken by applying a scale value of 20 at pixel level, being the second step carried out on the first

(*a*) (*b*)

Figure 3. Detailed (110 m × 135 m) multiresolution segmentation: (*a*) GeoEye-1 and (*b*) WV2.

segmentation level (bottom-up segmentation) by using a larger scale value of 70. The segmentation was always computed by taking into account the four equal-weighted bands corresponding to the pan-sharpened orthoimage.

In the case of WV2, the first segmentation at pixel level was carried out with a scale of 10.7, whereas the second bottom-up segmentation presented a scale value of 43. In this case, the eight equal-weighted bands corresponding to the WV2's pan-sharpened orthoimage were used to perform multiresolution segmentation.

3.2. Manual classification

As mentioned earlier, our main objective was to carry out a methodology to compare the capabilities of eight bands of WV2 and four bands of GeoEye-1 orthoimages using well-known and often used OBIA techniques in urban environments. Thus, to minimize the impact of the segmentation step on the final classification, two different ground references (manual classifications) for each segmentation were produced in the ArcGIS environment by careful visual inspection of separate data sources, including PAN and pan-sharpened orthoimages from GeoEye-1 and WV2, the orthoimage from Z/I Imaging's DMC, and the lidar-derived nDSM. In this sense, a reference map was manually produced by the visual inspection of each IO from GeoEye-1's segmentation. In the same way, another independent reference map was interpreted based on each IO from WV2's segmentation. This task was carried out and cross-validated by two interpreters, who considered ten target classes (Table 2). It is worth noting that the proposed accuracy assessment based on two independent ground reference maps, always exactly matching the segmented IOs, artificially removes the segmentation errors (i.e. extra pixels and lost pixels defined by Marpu et al. 2010).

For the GeoEye-1's segmentation, 1894 out of the initial 2723 IOs were visually indentified as meaningful objects (Table 2). A subset of 945 well-distributed IOs were selected to carry out the training phase, whereas the remaining 949 IOs, also well-distributed in the working area, were used for the validation phase. Note that for each class approximately 50% of the IOs belonged to the training subset, while the remaining 50% were applied to the validation subset. In the same way, and in the case of WV2's segmentation, 1759 out of 2720 IOs were manually classified (Table 3) and the training (876 IOs) and validation (883 IOs) subsets were assigned. The total surface areas occupied by IOs for each class

Table 2. IOs after GeoEye-1's segmentation and manual classification related to the target classes.

		Area (m^2)				
Class	No. IOs	Total area	Mean	σ	Validation IOs	Training IOs
Red Buildings	298	22217	74.55	51.62	149	149
White Buildings	558	17034.25	30.53	37.57	279	279
Grey Buildings	68	5279.5	77.64	57.11	34	34
Other Buildings	55	3086.25	56.11	34.04	28	27
Shadows	477	21600	45.28	63.59	239	238
Vegetation	194	17192.5	88.62	88.13	97	97
Bare Soil	93	15464	166.28	126.84	47	46
Roads	72	15720.5	218.34	161.99	36	36
Streets	71	7317.75	103.07	109.06	36	35
Swimming Pools	8	342.25	42.78	10.2	4	4

Table 3. IOs after WV2's segmentation and manual classification related to the target classes.

		Area (m^2)				
Class	No. IOs	Total area	Mean	σ	Validation IOs	Training IOs
Red Buildings	315	24209.5	76.86	50.81	158	157
White Buildings	315	10241.75	32.51	33.68	158	157
Grey Buildings	88	6528	74.18	61.95	44	44
Other Buildings	78	5116.75	65.60	49.57	39	39
Shadows	431	14404.5	33.42	32.48	216	215
Vegetation	159	13289	83.58	73.73	80	79
Bare Soil	147	18663.75	126.96	103.46	74	73
Roads	116	17489	150.77	96.21	58	58
Streets	97	8934.5	92.11	70.61	49	48
Swimming Pools	13	482.25	37.10	13.97	7	6

in the final ground references, as well as their mean and standard deviation (σ), are also shown in Tables 2 and 3.

The manually assigned reference maps based on previous segmentation results are shown in Figure 4, where the four classes related to buildings are presented as only one class, named Buildings. Although the number of non-classified IOs (spectrally mixed segments difficult to assign to an only target class; white pixels) in the reference maps is similar for both cases, the area occupied by these IOs is quite larger in the case of WV2 ground truth. This fact points out the under-segmentation problems. However, it should be noted that both IOs used as training samples and those employed to carry out the accuracy assessment were always meaningful objects.

3.3. Classifier and selection of training areas

A very well-known and traditional classifier like Nearest Neighbour (1-NN) was used in this work, although comparative studies have shown that classification by more sophisticated methods such as the support vector machine (SVM) approach can be more accurate than classical classifiers (Foody and Mathur 2004; Melgani and Bruzzone 2004). Despite SVM usually turning out to be more stable in high-dimensional feature spaces and needing only a small sample in training (e.g. Foody and Mathur 2006; Chen and Ho 2008), the classification accuracy of SVM models is largely dependent on the selection of the model's parameters (Chang and Lin 2011). From our point of view, this variability strongly supports the use of a parameter-free classifier like 1-NN for comparative purposes.

From GeoEye-1's segmentation, four different repetitions of 5%, 10%, 15%, and 20% IOs, respectively (16 training sets in total), were extracted from the training subset of 945 IOs. The training percentage was kept constant for each class, both according to the number of IOs and their mean area or size. For example, for GeoEye-1's test, the 298 IOs from the class Red Buildings yielded a mean area of 74.55 m^2 per object. Thus, for every different repetition chosen in this work, the training areas for the class Red Buildings had always a mean area of around 74 m^2. In the same way, another 16 different repetitions of training sets were selected from the training subset of 876 IOs in WV2's test. Table 4 depicts the number of IOs chosen for 1-NN classifier training in both cases.

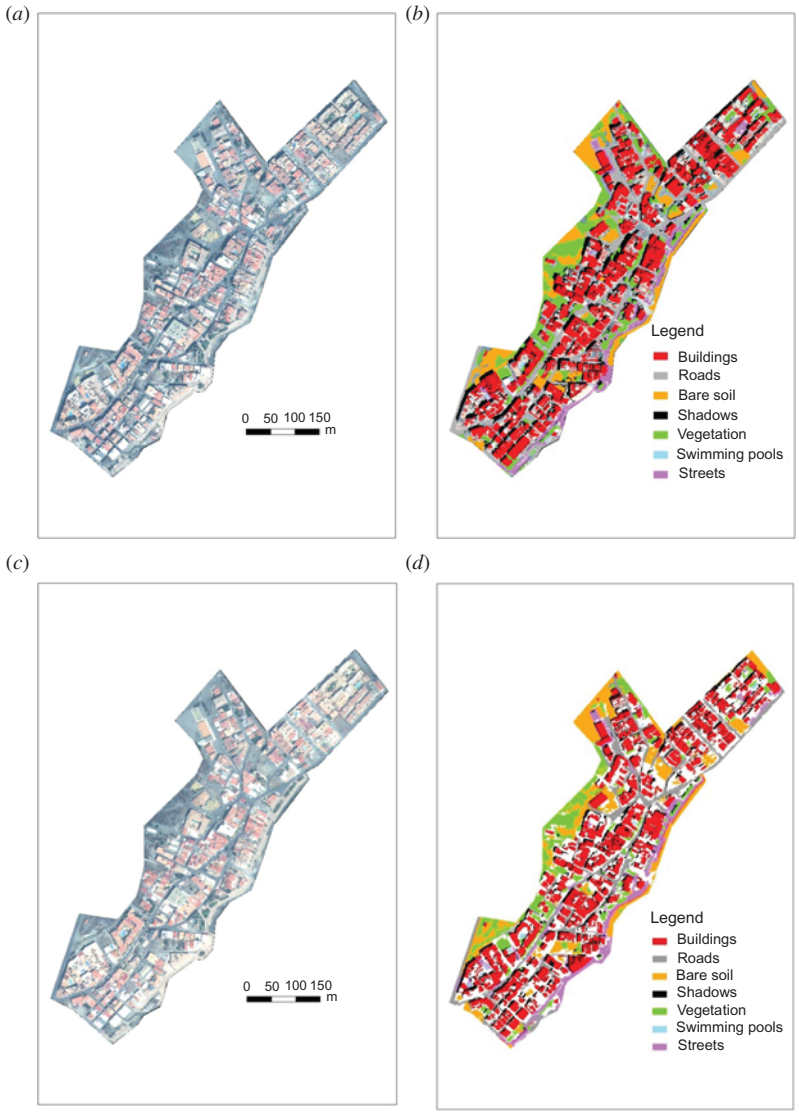

Figure 4. GeoEye-1 pan-sharpened orthoimage (*a*) and manual classifications from GeoEye-1 segmentation (*b*). WV2 pan-sharpened orthoimage (*c*) and manual classifications from WV2 segmentation (*d*).

3.4. Feature extraction and selection

The object features used to carry out the supervised classification are presented and described in Table 5. A total of 23 and 31 features, grouped into five categories, were tested for GeoEye-1 and WV2, respectively. More in-depth information about these features can be found in the Reference Book of Definiens eCognition Developer 8 (Definiens eCognition 2009). The mean and standard deviation values of the digital number for all of the pixels inside an IO for each band and the kind of orthoimages (pan-sharpened and PAN) were catalogued as Basic Spectral Information. Another three features (Normalized Difference Index for Blue, Green, and Red bands) extracted from the spectral information

Table 4. Number of IOs used for classifier training regarding each satellite and target class (four replicates for every training set).

	GeoEye-1				WV2			
Class	5% training	10% training	15% training	20% training	5% training	10% training	15% training	20% training
Red Buildings	15	30	45	60	16	32	48	63
White Buildings	28	56	84	112	16	32	48	63
Grey Buildings	4	7	11	14	5	9	14	18
Other Buildings	3	6	11	15	4	8	12	16
Shadows	24	48	72	96	22	44	65	87
Vegetation	10	20	30	39	8	16	24	32
Bare Soil	5	10	14	19	8	15	23	30
Roads	4	8	11	15	6	12	18	24
Streets	4	8	11	15	5	10	15	20
Swimming Pools	1	1	2	2	1	2	2	3

included in the pan-sharpened orthoimages (see Table 5) were grouped under the Band Index strategy. In the Texture category, only five texture features based on the grey-level co-occurrence matrix (GLCM) were considered among the 14 originally proposed by Haralick, Shanmugam, and Dinstein (1973) due to the strong correlation frequently reported between many of the features (e.g. Cossu 1988; Baraldi and Panniggiani 1995). They were always computed over the PAN image considering all the directions. The five selected features were contrast (con), entropy (ent), mean, standard deviation (stdv), and correlation (cor). The same subset of texture features had already been selected by Stumpf and Kerle (2011) working on a similar feature space for landslides mapping. Under the category Shape and Geometry, two geometric features based on shape (Rectangular Fit and Shape Index) and two based on polygons (Compactness and Number of Edges) were considered. Finally, last but not least important, one more feature corresponding to Elevation Data and containing the corresponding nDSM was also included.

Seven sets of features for both GeoEye-1 and WV2 were carried out, considering each of them as different strategies for classification.

(1) *Basic 1*. This set only included Basic Spectral Information (see Table 5). Ten feature types were considered for GeoEye-1's IOs, comprising the mean and standard deviation values computed for each band (Blue, Green, Red, NIR, and PAN). In this sense, 18 feature types were computed for WV2 (Coastal, Blue, Green, Yellow, Red, Red Edge, NIR1, NIR2, and PAN). An alternative strategy named Basic 2 (10 feature types) was tested only for WV2. In this case, the information corresponding to the newest bands (i.e. Coastal, Yellow, Red Edge, and NIR2) was not considered.
(2) *Elevation*. It was composed of the Basic 1 set plus the feature mean elevation for each IO coming from the nDSM (11 and 19 feature types for GeoEye-1 and WV2, respectively).
(3) *NDIs*. It was composed of the Basic 1 set plus the three Normalized Difference Indexes (13 and 21 feature types for GeoEye-1 and WV2, respectively).
(4) *Elevation + NDIs*. Basic 1 set plus NDIs and nDSM (14 and 22 feature types for GeoEye-1 and WV2, respectively).

Table 5. IO features used in the classification phase.

	Tested features	Applied to	Description
Basic spectral information	Coastal	Only WV2	Mean and standard deviation of pan-sharpened Coastal band
	Blue	Both	Mean and standard deviation of pan-sharpened Blue band
	Green	Both	Mean and standard deviation of pan-sharpened Green band
	Yellow	Only WV2	Mean and standard deviation of pan-sharpened Yellow band
	Red	Both	Mean and standard deviation of pan-sharpened Red band
	Red Edge	Only WV2	Mean and standard deviation of pan-sharpened Red Edge band
	NIR1	Both	Mean and standard deviation of pan-sharpened NIR1 band
	NIR2	Only WV2	Mean and standard deviation of pan-sharpened NIR2 band
	PAN	Both	Mean and standard deviation of PAN band
Band index	NDBI	Both, NIR1 for WV2	Normalized difference of Blue band Index (Mean NIR – Mean Blue)/(Mean NIR + Mean Blue)
	NDGI	Both, NIR1 for WV2	Normalized difference of Green band Index (Mean NIR – Mean Green)/(Mean NIR + Mean Green)
	NDVI	Both, NIR1 for WV2	Normalized difference vegetation index (Mean NIR – Mean Red)/(Mean NIR + Mean Red)
Texture	$GLMC_{con}$	Both	GLCM contrast sum of all directions from PAN
	$GLMC_{ent}$	Both	GLCM entropy sum of all directions from PAN
	$GLMC_{mean}$	Both	GLCM mean sum of all directions from PAN
	$GLMC_{stdv}$	Both	GLCM standard deviation sum of all directions from PAN
	$GLMC_{cor}$	Both	GLCM correlation sum of all directions from PAN
Shape and geometry	Rectangular fit	Both	It describes how well an image object fits into a rectangle
	Shape Index	Both	The border length of the IO divided by four times the square root of its area
	Compactness	Both	The ratio of the area of a polygon to the area of a circle with the same perimeter
	No. Edges	Both	The number of edges that form the polygon
Elevation data	nDSM	Both	Lidar-derived normalized digital surface model

Note: NIR1, Near Infrared-1; NIR2, Near Infrared-2.

(5) *Texture*. Basic 1 set plus texture feature types (15 and 23 features for GeoEye-1 and WV2, respectively).
(6) *Geometry*. Basic 1 set plus shape and geometric feature types (14 and 22 features for GeoEye-1 and WV2, respectively).
(7) *All*. All the features presented in Table 5 (23 and 31 features for GeoEye-1 and WV2, respectively).

3.5. Classification and accuracy assessment

The seven aforementioned strategies (eight for WV2 when considering Basic 2) were run by applying 1-NN classifier to extract the target classes. Bearing in mind that there were 16 sets of training samples for each sensor, 240 different classification projects were carried out in eCognition (i.e. 112 and 128 for GeoEye-1 and WV2, respectively). In all of them, the accuracy assessment was computed by means of an error matrix based on IOs. For GeoEye-1's classification projects, classification accuracy was finally assessed on a validation set always composed of the same 949 IOs (Table 2), whereas the 883 validation segments (Table 3) were kept constant for the WV2's projects.

It is noteworthy that the four classes related to buildings (i.e. Red, White, Grey, and Other buildings) were grouped in only one class named Buildings before computing the accuracy indexes explained later. Note that the class Buildings was the more extended class along the working area, presenting a total area percentage over the manual classification surface of more than 38% both for GeoEye-1 and WV2 (see Tables 2 and 3). On the other hand, Streets, and especially Swimming pools, were the two classes covering less extension along the working area.

User's accuracy (UA), producer's accuracy (PA), and overall accuracy (OA), together with the overall and per class (Buildings, Shadows, Vegetation, Bare Soil, Roads, Streets, and Swimming Pools) kappa index (Congalton 1991; Congalton and Green 2009), were the accuracy values (based on error matrix) computed in this work. UA details errors of commission (i.e. when an object is committed to an incorrect class), whereas PA details the errors of omission (i.e. when a segment is incorrectly classified into another category and so it is omitted from its correct class). Furthermore, OA is the number of correctly classified objects divided by the total number of objects. Finally, the kappa statistic incorporates the off-diagonal observations of the rows and columns, as well as the diagonal values of the error matrix, to give a more robust accuracy assessment than OA measures.

3.6. Statistical analysis

In order to study the influence of the analysed factors (i.e. type of sensor, different strategies, and size of training samples) on the final classification accuracy, several analysis of variance (ANOVA) tests for three factors (univariate general linear model) or only for one of them (one-way ANOVA) were carried out by means of a factorial model with four repetitions (Snedecor and Cochran 1980). The observed variables were OA, PA, UA, and overall and per class kappa index of agreement (KIA and KIA_{class}). When the results of the ANOVA test turned out to be significant ($p < 0.05$), the separation of means was carried out using the Duncan's multiple range test at 95% confidence level. Furthermore, the Kolmogorov–Smirnov test was used as a goodness of fit to a standard normal distribution.

4. Results

4.1. Differences between tested images

First of all it is important to highlight that clear visual differences can be appreciated between the final pan-sharpened orthoimages from GeoEye-1 (Figure 2(*a*)) and WV2 (Figure 2(*b*)). The GeoEye-1 image presented a higher off-nadir angle (20.6° off-nadir) than the one corresponding to the WV2 image (10° off-nadir). This greater tilt of GeoEye-1 caused an increase in image-projected building facades (usually white), resulting in a higher number of IOs classified as White Buildings. Moreover, sun elevation turned out to be quite different between GeoEye-1 and WV2, taking values of 48° and 64°, respectively. It provoked an important increase of shadows in the GeoEye-1 orthoimage with respect to the WV2 orthoimage. Also, vegetation looked quite different and, casually, the number of objects (e.g. vehicles) situated along streets was larger in the case of the WV2 image. Finally, the WV2 orthoimage looked a little bit fuzzy, being blurrier and hence showing a lesser contrast than the GeoEye-1 orthoimage. In this sense, perhaps as a result of different post-processing carried out by the imaging companies, a notable compression of the histogram for each band was observed on both the PAN and MS products contained in the raw WV2 image (Figure 5), especially for PAN and blue bands. This behaviour has also been observed in a WV2 ORS2A bundle image taken in July 2011 used by Aguilar, Saldaña, and Aguilar (2012b) over the same working area. Also, this effect can be clearly appreciated, as compared with the GeoEye-1 images, in Figure 2 of the recent work published by Agugiaro, Poli, and Remondino (2012) over Trento testfield (Italy).

Although the observed blurring effect might be due to the operational aspects of image acquisition such as sensor viewing angle, sun acquisition angle, and atmospheric conditions (Poli, Angiuli, and Remondino 2010), we hypothesized that differences in sampling rate, footprint speed, and the specific radiometric characteristics of both sensor systems might be the main factor responsible for the more accentuated blurring effect found in the WV2 image. Thus, the results obtained from this work should be carefully managed and contrasted with further studies.

4.2. Overall classification accuracy

The observed variables related to the classification accuracy assessment fit well to a normal distribution (Kolmogorov–Smirnov test) except for the accuracy values corresponding to the class Swimming pools. This non-normal distribution was due to the small number of segments manually classified as Swimming pools for both GeoEye-1 (8) and WV2 (13) orthoimages. In this sense, the results obtained from the class Swimming pools will be ignored in this work.

The first global ANOVA test was performed considering the effects on the overall classification accuracy, measured as OA and KIA, of the three analysed factors. From this analysis, it can be highlighted that the OA and KIA values showed a similar behaviour.

In short, three main factors turned out to be significant ($p < 0.05$), the type of satellite being the most important source of variation, followed by the applied features vector, and finally the percentage of training. The cross-interaction between type of sensor and features vector was also significant, although its corresponding F-test value was lesser than the one computed for the three main factors.

Table 6 shows the comparison of OA and KIA mean values according to the first global ANOVA test. With regard to the set of features tested, note that geometry and texture features had little influence on the OA and KIA values, whereas both nDSM and NDIs added valuable information to the Basic 1 strategy. Concerning training size, Table 6 confirms that

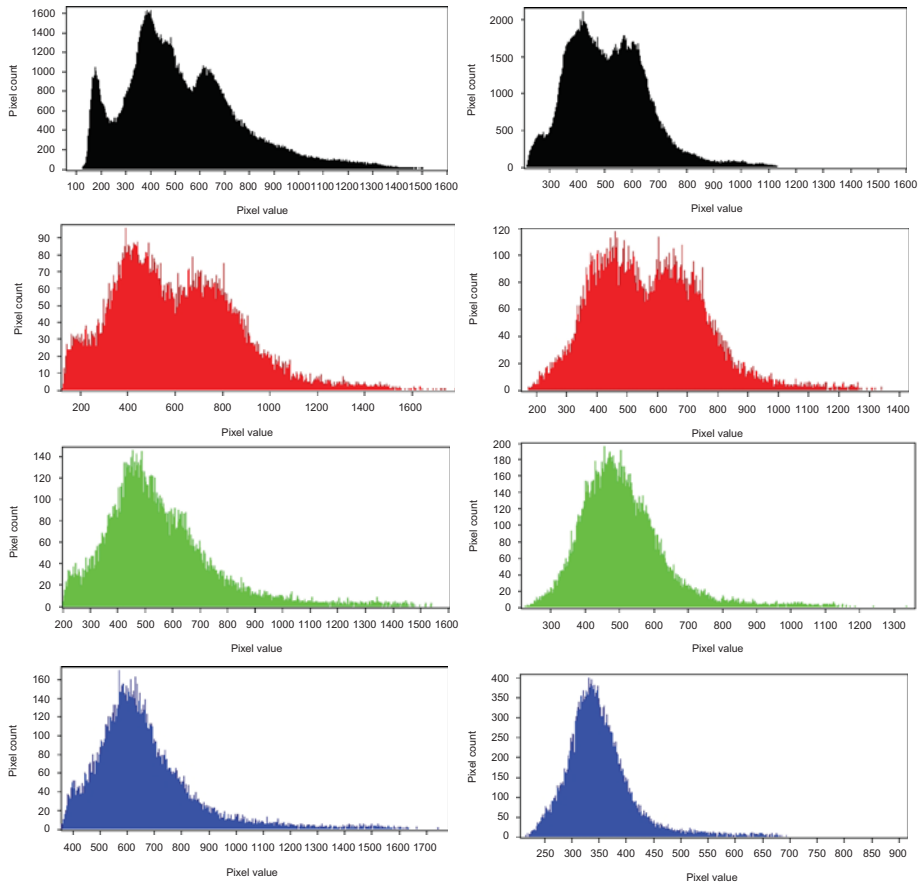

Figure 5. Histograms of digital numbers from PAN, MS Red, MS Green, and MS Blue bands of the original GeoEye-1 (left) and WV2 (right) images used in this work.

the larger the number of training IOs the better. Finally, there was a significant difference between the overall classification accuracy (OA and KIA) provided by our GeoEye-1 and WV2 orthoimages. The results regarding overall classification accuracy for each satellite will be explored in depth in the following sections by focusing on each one of the sets of features and training areas tested.

4.2.1. Sets of features

Table 7 shows the independent classification accuracy assessment for each sensor considering only the factor set of features (one-way ANOVA). In general, for both satellites, the best accuracies in terms of OA and KIA were achieved just when the feature nDSM was included (i.e. Elevation, Elevation + NDIs, and All). Another similarity was that the Geometry set did not work well in any case. On the other hand, bearing in mind that all findings are only valid for the special conditions of both used images here, higher OA and KIA values were attained in the case of GeoEye-1. Moreover, as compared to the Basic 1 strategy, the texture features significantly improved both OA and KIA for the WV2 test, whereas any statistical effect was detected in the case of GeoEye-1. Focusing on GeoEye-1,

Table 6. Comparison of mean values for the global ANOVA of overall accuracy (OA) and kappa (KIA) for the three main factors tested.

Factors	Disaggregated factors	OA (%)	KIA
Sets of Features	Geometry	75.0[f]	0.633[f]
	Basic 1	76.8[e]	0.663[e]
	Texture	77.9[d]	0.679[d]
	NDIs	79.5[c]	0.703[c]
	Elevation	80.8[b]	0.722[b]
	Elevation + NDIs	82.9[a]	0.753[a]
	All	83.6[a]	0.760[a]
Training	5%	77.9[d]	0.677[d]
	10%	78.8[c]	0.693[c]
	15%	80.4[b]	0.714[b]
	20%	81.0[a]	0.724[a]
Satellite	WV2	75.7[b]	0.659[b]
	GeoEye-1	83.4[a]	0.743[a]

Note: For each factor tested, values in the same column followed by different superscript letters indicate significant differences at a significance level, $p < 0.05$.

Table 7. Comparison of mean values of overall accuracy (OA) and kappa (KIA) for the sets of features or strategies and sensor tested.

	GeoEye-1		WV2	
Strategies	OA (%)	KIA	OA (%)	KIA
Basic 2	–	–	71.3[d]	0.598[d]
Geometry	78.3[d]	0.665[d]	71.7[d]	0.601[d]
Basic 1	81.5[c]	0.717[c]	72.1[d]	0.609[d]
Texture	81.0[c]	0.707[c]	74.9[c]	0.652[c]
NDIs	84.9[b]	0.769[b]	74.1[c]	0.638[c]
Elevation	84.3[b]	0.761[b]	77.4[b]	0.683[b]
Elevation + NDIs	87.2[a]	0.806[a]	78.6[b]	0.701[b]
All	86.4[a]	0.791[a]	80.7[a]	0.729[a]

Note: Values in the same column followed by different superscript letters indicate significant differences at a significance level, $p < 0.05$.

the best performance sets of features ($p < 0.05$) were both All and Elevation + NDIs sets, yielding an OA value close to 87% (i.e. both values are followed by the letter 'a', which is not repeated in any other value of the same column). However, looking at the WV2 case, the best and significant ($p < 0.05$) accuracy was attained by using the set of features All, achieving OA values in the range of 80%.

It is worth noting that, under our experimental conditions, the strategy named Basic 2, where only the MS traditional four bands of WV2 were considered (i.e. Blue, Green, Red, and NIR1), produced the same accuracy results as Basic 1 (i.e. no statistical differences).

4.2.2. Training size

The influence of training size was studied here for each satellite. Table 8 shows the comparison of mean results from the one-way ANOVA analysis carried out for the overall classification accuracy (OA and KIA). The rule of 'the larger the better' is also true here, since OA and KIA always improved when a larger percentage of training was applied.

Table 8. Comparison of mean values of overall accuracy (OA) and kappa (KIA) for the training percentage of samples and sensor tested.

	GeoEye-1		WV2	
Size of training (%)	OA (%)	KIA	OA (%)	KIA
5	81.7[b]	0.719[b]	73.6[b]	0.627[b]
10	83.1[a,b]	0.741[a,b]	74.0[b]	0.6373[b]
15	84.1[a]	0.756[a]	76.2[a]	0.665[a]
20	84.7[a]	0.765[a]	76.7[a]	0.675[a]

Note: Values in the same column followed by different superscript letters indicate significant differences at a significance level, $p < 0.05$.

However, in statistical terms, the most efficient choice in the case of GeoEye-1 would be about 10% of training samples (this size is not statistically different to 15% or 20%), whereas 15% of training would be the best option for WV2. This fact could be related to the different number of available bands in each sensor. In other words, a larger size of training areas is needed for a higher number of bands (Piper 1992; Mather 2004). As will be outlined later, training size is very dependent on the target class.

4.3. Buildings

The class Buildings is usually the most important in urban environments, in our case representing more than 38% of the manual classified area. According to Table 9, the most relevant feature for classifying Buildings from both GeoEye-1 and WV2 turned out to be the nDSM from laser scanning. This fact has already been reported by many authors such as Hermosilla et al. (2011), Awrangjeb, Ravanbakhsh, and Fraser (2010), Turker and San (2010), or Longbotham et al. (2012). Paying attention to our GeoEye-1 image, the best strategies were Elevation + NDIs and All features, achieving consistently high-accuracy values in the range of 95% for both PA and UA. On the other hand, Elevation, Elevation + NDIs, and All features were the best performance strategies (without significant differences between them) for WV2 orthoimages, attaining PA and OA values close to 91%.

Table 9. Comparison of mean values for the class Buildings of producer's accuracy (PA), user's accuracy (UA), and kappa ($KIA_{buildings}$) for the sets of features or strategies and sensor tested.

	GeoEye-1			WV2		
Strategies	PA (%)	UA (%)	$KIA_{buildings}$	PA (%)	UA (%)	$KIA_{buildings}$
Basic 2	–	–	–	85.5[b]	83.5[d]	0.731[b]
Geometry	89.4[d]	88.2[d]	0.777[d]	84.7[b]	83.6[d]	0.720[b]
Basic1	89.1[d]	89.7[c]	0.777[d]	83.9[b]	84.5[c,d]	0.705[b]
Texture	91.0[c]	89.6[c]	0.811[c]	85.0[b]	85.7[b]	0.734[b]
NDIs	91.4[c]	91.0[b]	0.822[c]	85.0[b]	85.4[b,c]	0.728[b]
Elevation	94.0[b]	94.6[a]	0.877[b]	91.1[a]	91.7[a]	0.839[a]
Elevation + NDIs	95.3[a]	94.9[a]	0.903[a]	90.8[a]	91.2[a]	0.834[a]
All	95.8[a]	94.9[a]	0.911[a]	92.1[a]	91.6[a]	0.856[a]

Note: Values in the same column followed by different superscript letters indicate significant differences at a significance level, $p < 0.05$.

In the case of the class Buildings, the influence of training size was not significant regarding classification accuracy indexes (PA, UA, and $KIA_{buildings}$), though a slight trend was observed to improve them by increasing the percentage of training samples (data not presented). It is important to highlight that only 5% of training percentage for Buildings already included more than 40 segments for both GeoEye-1 and WV2.

4.4. Shadows

Classifying the class Shadows was extremely challenging since shadows vary spectrally based on the features that are casting them. In addition, the spectral characteristics of Shadows can be either similar to other surface materials such as water (Sawaya et al. 2003) and dark or grey roof (Bhaskaran, Paramananda, and Ramnarayan 2010). Furthermore, shadows are particularly significant in high-spatial resolution imagery in urban environments, where elevation varies dramatically across short distances (Dare 2005; Yuan 2008). The introduction of the class Shadows is necessary to prevent shaded pixels in the urban area from being classified as water or dark roof building (Shackelford and Davis 2003). Table 10 shows that the tested GLMC texture features were the best choice for their classification by means of 1-NN from both GeoEye-1 and WV2 orthoimages. Texture features are a potentially powerful method for detecting Shadows as they are highly distinctive, do not depend on colours, and are robust to illumination changes. In this way, Su et al. (2008) improved the Shadows classification in urban areas by using textural features based on GLMC as additional information bands. However, GLMC-based features require high computation time.

Summing up, the best results in terms of PA, UA, and $KIA_{shadows}$ were attained by using Texture or All strategies, achieving an accuracy of around 90% (PA and UA) for GeoEye-1 and slightly lower values for WV2 orthoimages. In this case, the statistically significant worst results for WV2 were obtained when using the set of features named Basic 2.

In the same way as Buildings, Shadows classification accuracy did not depend on the training size used (data not presented). Note that in the case of class Shadows, 5% of training samples involved 24 and 22 IOs for GeoEye-1 and WV2, respectively. In this way, at least 15 sample plots for each training class were used by Lu, Hetrick, and Moran (2010) working with QuickBird imagery.

Table 10. Comparison of mean values for the class Shadows of producer's accuracy (PA), user's accuracy (UA), and kappa ($KIA_{shadows}$) for the sets of features or strategies and sensor tested.

	GeoEye-1			WV2		
Strategies	PA (%)	UA (%)	$KIA_{shadows}$	PA (%)	UA (%)	$KIA_{shadows}$
Basic 2	—	—	—	71.6^{c}	79.7^{e}	0.636^{c}
Basic 1	92.7^{a}	$88.8^{a,b}$	0.901^{a}	84.2^{b}	$81.8^{c,d,e}$	0.779^{b}
Geometry	88.6^{c}	86.4^{c}	0.846^{c}	85.1^{b}	$80.9^{d,e}$	0.799^{b}
NDIs	$89.9^{b,c}$	90.4^{a}	$0.865^{b,c}$	83.7^{b}	$83.0^{c,d}$	0.783^{b}
Elevation	$90.3^{b,c}$	86.4^{c}	$0.869^{b,c}$	84.0^{b}	$82.9^{c,d}$	0.788^{b}
Elevation + NDIs	88.7^{c}	$89.1^{a,b}$	0.850^{c}	83.4^{b}	83.5^{b}	0.780^{b}
Texture	92.7^{a}	$88.9^{a,b}$	0.902^{a}	89.5^{b}	$85.6^{a,b}$	0.857^{a}
All	$91.1^{a,b}$	88.7^{b}	$0.881^{a,b}$	89.7^{a}	87.0^{a}	0.863^{a}

Note: Values in the same column followed by different superscript letters indicate significant differences at a significance level, $p < 0.05$.

4.5. Vegetation

Regarding the class named Vegetation (Table 11), the overall results for GeoEye-1 and WV2 were very heterogeneous. However, mainly focusing on $KIA_{vegetation}$, the best performance sets of features were those containing the NDIs features. It is noteworthy that one of the three indexes included in the NDIs set is the well-known normalized difference vegetation index (NDVI). In fact, Vinciková et al. (2010) reported that NDVI is one of the most commonly used vegetation indices in remote-sensing applications. Moreover, Zerbe and Liew (2004) pointed out that the normalized difference of blue band index (NDBI) could help to distinguish Vegetation class. In another way, authors such as Haala and Brenner (1999) demonstrated that the use of lidar data improved tree detection, beside buildings, in an urban area. On the basis of these above data, Elevation + NDIs set could be considered as the optimal choice for the detection of class Vegetation.

For class Vegetation, as opposite to Buildings and Shadows classes, the classification accuracy ($KIA_{vegetation}$) depended on the percentage of training samples used (data not presented). The more efficient training size, in statistical terms, looked very similar to the one shown in Table 8 from KIA figures. It meant an optimum training percentage of 10% in GeoEye-1 case (20 training IOs) and 15% for WV2 (24 training IOs).

4.6. Roads, Bare Soil, and Streets

The results of the three remaining classes (i.e. Roads, Bare Soil, and Streets) are summarized in Table 12 by means of KIA per class values. The Bare Soil class achieved very poor accuracy results, with the best PA and UA values for both GeoEye-1 and WV2 ranging from 45.5% to 53.3%, respectively. It could be due to the high heterogeneity of this class, which included agricultural soils, non-asphalted roads, building lots, and even beaches. The same can be applied to the class Streets that groups together pavements, concrete streets, or sport courts. Misclassification problems were detected between Buildings and Streets owing to spectral similarities regarding materials covering these surfaces. However, many of them were corrected when nDSM was included in the classification project (Table 12). Regarding the class Streets, the best PA and UA values (PA = 59.9% and UA = 65.2% for GeoEye-1, PA = 50% and PA = 50.7% for WV2) were always attained by using Elevation + NDIs set.

Table 11. Comparison of mean values for the class Vegetation of producer's accuracy (PA), user's accuracy (UA), and kappa ($KIA_{vegetation}$) for the sets of features or strategies and sensor tested.

	GeoEye-1			WV2		
Strategies	PA (%)	UA (%)	$KIA_{vegetation}$	PA (%)	UA (%)	$KIA_{vegetation}$
Basic 2	–	–	–	$68.8^{a,b,c}$	68.6^{b}	$0.658^{a,b}$
Basic 1	71.2^{c}	83.8^{b}	0.685^{c}	$64.1^{b,c}$	69.5^{b}	$0.615^{b,c}$
Geometry	54.4^{e}	63.1^{e}	0.501^{e}	54.2^{d}	63.8^{c}	0.504^{d}
Texture	63.3^{d}	71.4^{d}	0.596^{d}	63.4^{c}	$71.3^{a,b}$	0.588^{c}
Elevation	72.3^{c}	79.6^{c}	0.696^{c}	$66.4^{a,b,c}$	69.4^{b}	$0.633^{a,b,c}$
NDIs	86.1^{a}	90.1^{a}	0.846^{a}	$69.7^{a,b,c}$	$70.9^{a,b}$	$0.668^{a,b}$
Elevation + NDIs	86.7^{a}	88.8^{a}	0.853^{a}	72.0^{a}	$71.7^{a,b}$	0.693^{a}
All	79.6^{b}	$86.1^{a,b}$	0.776^{b}	$70.5^{a,b}$	75.7^{a}	$0.678^{a,b}$

Note: Values in the same column followed by different superscript letters indicate significant differences at a significance level, $p < 0.05$.

Table 12. Comparison of mean values for the classes Roads, Bare Soil, and Streets of producer's accuracy (PA), user's accuracy (UA), and kappa per class for the sets of features or strategies and sensor tested.

	GeoEye-1			WV2		
Strategies	KIA_{roads}	KIA_{bare}	$KIA_{streets}$	KIA_{roads}	KIA_{bare}	$KIA_{streets}$
Basic 2	–	–	–	0.560^{a}	0.209^{d}	$0.293^{b,c}$
Basic 1	0.368^{d}	$0.366^{b,c}$	$0.315^{c,d}$	$0.422^{d,e}$	0.279^{c}	$0.288^{b,c}$
Geometry	0.350^{d}	$0.339^{b,c}$	0.209^{e}	0.397^{e}	0.290^{c}	0.244^{c}
Texture	0.353^{d}	0.305^{c}	$0.249^{d,e}$	$0.453^{c,d,e}$	$0.330^{b,c}$	0.321^{b}
Elevation	0.358^{d}	$0.407^{a,b}$	0.588^{a}	$0.452^{c,d,e}$	0.347^{b}	0.469^{a}
NDIs	0.550^{a}	0.453^{a}	$0.372^{b,c}$	$0.483^{b,c,d}$	0.287^{c}	0.324^{b}
Elevation + NDIs	0.488^{b}	0.473^{a}	0.637^{a}	$0.519^{b,c}$	$0.371^{a,b}$	0.481^{a}
All	0.428^{c}	0.468^{a}	0.444^{b}	$0.541^{a,b}$	0.407^{a}	0.438^{a}

Note: Values in the same column followed by different superscript letters indicate significant differences at a significance level, $p < 0.05$.

Regarding the class Roads (Table 12), the best accuracy for GeoEye-1 was achieved by using the set of feature NDIs. It is noted that this strategy included NDBI, which had been already pointed out by Dinis et al. (2010) to discriminate between bare soil and roads from a QuickBird satellite orthoimage. In addition, the set of features named All and, surprisingly, Basic 2 reached the best accuracy results in the case of the WV2 image. The PA and UA values for GeoEye-1 and WV2 ranged from 52.6% to 62.6%.

From a statistical point of view, our results indicate that the more efficient WV2 training size would be close to 10% (i.e. 15 IOs corresponding to Bare Soil and 10 to Streets). In the case of GeoEye-1, the best accuracy results were attained by using 10 Bare Soil IOs (10% of training percentage) and 15 Streets IOs (20% of training percentage). In the case of class Roads, again 10% was recommended as the more adequate training size, meaning 8 and 12 Roads IOs for GeoEye-1 and WV2, respectively.

5. Discussion

Bearing in mind the aforementioned differences between the images tested in this work and the typical heterogeneity within urban landscape classes, especially in VHR satellite images, the best overall classification accuracy attained here by using the tested GeoEye-1 four bands pan-sharpened orthoimage, as compared with the results offered by the WV2 eight bands pan-sharpened orthoimage, should not be taken as a conclusive result. In fact, the authors, once this differentiated performance has been detected, are trying to find out whether this is due to satellite image quality differences (blurrier images in the case of WV2). To the knowledge of the authors, this is the first publication that points out the possible radiometric differences between both VHR satellites. Thus, it is extremely important to take this research line on to investigate whether the blurring effect shown by the tested WV2 image might actually affect the final classification accuracy results. Yet there are some other causes that could explain the obtained results and so they would have to be tested through further works. Among others, the following hypotheses could be stressed.

(1) The different off-nadir angles and sun elevations between the original images from WV2 and GeoEye-1 used through this work. It caused substantial differences mainly in the area covered by Shadows and White Buildings (presence of facade

effect due to high off-nadir angles) classes. To avoid this undesirable effect, which could pollute somehow the final comparative accuracy results, a new test involving two images from WV2 and GeoEye-1 with very similar off-nadir and sun elevation should be carried out.

(2) The multiresolution segmentation was carried out for attaining, approximately, the same number of IOs from WV2 and GeoEye-1 orthoimages. In this way, it is very probable that segmentation results were not optimal in both cases and so under-segmentation error could adulterate the final classification accuracy results (Liu and Xia 2010). Note that WV2 segmentation has been somehow forced to fit the goal of obtaining approximately the same number of IOs for each tested image. In this sense, including thematic layers of buildings and/or nDSM in the segmentation approach could help to achieve a better match between the considered classes and the extracted IOs to make them spectrally purer. However, the goal in this article was not headed up to compare the best possible results from an OBIA process (obviously the best segmentation would be necessary to produce the best final mapping results via object-based classification) but comparing exclusively the relative classification accuracy regarding pan-sharpened orthoimages from GeoEye-1 and WV2 VHR satellites. It has been faced through the manual classification of two object-based ground truths (one extracted from each previously segmented satellite orthoimage) to carry out an independent accuracy assessment for every tested classification project based on its corresponding ground truth.

(3) Classifications algorithms frequently used in object-based classification, as is the case of 1-NN, do not perform well on a high-dimensional feature space due to problems related to feature correlation (the widely known curse of dimensionality). In this sense, for WV2's Basic 1 strategy, 16 features were used, whereas only eight were involved in the same strategy for GeoEye-1. However, bearing in mind that there were no significant differences between the overall results attained for WV2 from both Basic 1 and Basic 2 strategies (Table 7), this hypothesis could be considered as clearly weak. As a further work, to avoid as much as possible this disturbing effect, we propose to apply SVM-based classification algorithms that have previously shown a good performance in dealing with a large number of features (Melgani and Bruzzone 2004).

6. Conclusions

The overall accuracy assessment tests carried out in this study showed significant differences ($p < 0.05$) regarding object-based classification accuracy in urban environments gathered from both GeoEye-1 and WV2 pan-sharpened orthoimages. The overall accuracies attained from GeoEye-1 (four bands) orthoimages were always better than the computed ones from WV2 (eight bands) for every set of features tested. Disaggregating the results by class, the classification accuracy from GeoEye-1 significantly improved the one attained from WV2 in the cases of Buildings, Shadows, Vegetation, and Bare Soil classes. However, mainly due to the differences found between the images tested, these results should be contrasted by means of further works. In fact, hypotheses related to the disturbance effects due to the different off-nadir angles and sun elevations between the original images used in this work, the segmentation approach and likely under-segmentation problems, the high-dimensional feature space used in some tested classification strategies, and the differences observed between the radiometric distribution and visual appearance of PAN and MS images from both sensors (blurrier images in the case of WV2), have been already raised in Section 5.

Regarding the set of features applied to carry out the classification, the best overall accuracy from GeoEye-1 orthoimages (87.2%) was attained when the Elevation + NDIs set of features was used, whereas All features had to be taken into account for attaining the best overall accuracy in the case of WV2 orthoimages (80.7%). It is noteworthy that the All features strategy included the five texture features based on GLCM, being much more time consuming in terms of CPU requirements. On the other hand, the shape and geometric features (Geometry) did not contribute to improving the classification attained by the basic spectral features set (Basic 1). Perhaps its contribution could be upgraded by using better initial segmentations and so better defined IOs geometries. In the particular case of WV2, there were no significant differences between the overall results offered by Basic 1 (eight bands) and Basic 2 (using the four equivalent bands to GeoEye-1) strategies.

In general, nDSM was the most important feature for detecting Buildings, as it has been already reported by many authors working on different sources of images, such as Ikonos, WV2, or digital aerial images. The inclusion of NDVI, together with NDBI within the NDIs strategy, significantly improved the classification of Vegetation and Roads classes. In the same way, texture features added valuable information for detecting the Shadows class.

Finally, the other significant factor in the classification accuracy results was training size. For the main classes (Buildings, Shadows, and Vegetation), around 20 training IOs were needed for attaining the best results, whereas for Roads, Bare Soil, and Streets a number of IOs ranging from 8 to 15 were enough. It must be underlined that the IOs from the last three classes were larger than those from the first three (i.e. each IO covered a larger area on average, so occupying more image pixels). In this way, 20 training IOs would be enough for attaining good accuracies from using object-based supervised classification by applying 1-NN classifier.

Acknowledgements

This work was supported by the Spanish Ministry for Science and Innovation (Spanish Government) and the European Union (FEDER funds) under Grant Reference CTM2010-16573. The authors also appreciate the support from Andalusia Regional Government, Spain, through the Excellence Research Project RNM-3575. The authors are very grateful to Rubén Vicente Ferrón (University of Almeria) and Dr Antonio Fernández Álvarez (University of Vigo) for their contribution to this work.

References

Addink, E. A., S. M. de Jong, and E. J. Pebesma. 2007. "The Importance of Scale in Object-Based Mapping of Vegetation Parameters with Hyperspectral Imagery." *Photogrammetric Engineering and Remote Sensing* 73: 905–12.

Agugiaro, G., D. Poli, and F. Remondino. 2012. "Testfield Trento: Geometric Evaluation of Very High Resolution Satellite Imagery." In *Proceeding of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (Vol. XXXIX-B8), XXII ISPRS Congress*, August 25–September 1, Melbourne, Australia.

Aguilar, M. A., F. J. Aguilar, M. M. Saldaña, and I. Fernández. 2012a. "Geopositioning Accuracy Assessment of GeoEYE-1 Panchromatic and Multispectral Imagery." *Photogrammetric Engineering and Remote Sensing* 78: 247–57.

Aguilar, M. A., M. M. Saldaña, and F. J. Aguilar. 2012b. "Assessing Geometric Accuracy of the Orthorectification Process from GeoEye-1 and WorldView-2 Panchromatic Images." *International Journal of Applied Earth Observation and Geoinformation*, doi:10.1016/j.jag.2012.06.004.

Albrecht, F., S. Lang, and D. Hölbling. 2010. "Spatial Accuracy Assessment of Object Boundaries for Object-Based Image Analysis." In *Proceeding of the International Archives* of the *Photogrammetry, Remote Sensing* and *Spatial Information Sciences (Vol. XXXVIII-4/C7), Geographic Object-Based Image Analysis (GEOBIA 2010)*, June 29–July 2, Ghent, Belgium.

Awrangjeb, M., M. Ravanbakhsh, and C. S. Fraser. 2010. "Automatic Detection of Residential Buildings Using LIDAR Data and Multispectral Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 457–67.

Baatz, M., and M. Schäpe. 2000. "Multiresolution Segmentation – An Optimization Approach for High Quality Multi-Scale Image Segmentation." In *Angewandte Geographische Informations-Verarbeitung XII*, edited by J. Strobl, T. Blaschke, and G. Griesebner, 12–23. Karlsruhe: Wichmann Verlag.

Baraldi, A., and F. Panniggiani. 1995. "An Investigation of the Textural Characteristics Associated with Gray Level Cooccurrence Matrix Statistical Parameters." *IEEE Transactions on Geoscience and Remote Sensing* 33: 293–304.

Bhaskaran, S., S. Paramananda, and M. Ramnarayan. 2010. "Per-Pixel and Object-Oriented Classification Methods for Mapping Urban Features Using Ikonos Satellite Data." *Applied Geography* 30: 650–65.

Blaschke, T. 2010. "Object Based Image Analysis for Remote Sensing." *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 2–16.

Carleer, A. P., and E. Wolff. 2006. "Urban Land Cover Multi-Level Region-Based Classification of VHR Data by Selecting Relevant Features." *International Journal of Remote Sensing* 27: 1035–51.

Chang, C. C., and C. J. Lin. 2011. "LIBSVM: A Library for Support Vector Machines." *ACM Transactions on Intelligent Systems and Technology* 2: 1–27.

Chen, C. H., and P. G. P. Ho. 2008. "Statistical Pattern Recognition in Remote Sensing." *Pattern Recognition* 41: 2731–41.

Congalton, R. G. 1991. "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data." *Remote Sensing of Environment* 37: 35–46.

Congalton, R. G., and K. Green. 2009. *Assessing the Accuracy of Remotely Sensed Data. Principles and Practices*. 2nd ed. Boca Raton, FL: CRS Press/Taylor & Francis.

Cossu, R. 1988. "Segmentation by Means of Textural Analysis." *Pixel* 1: 21–4.

Dare, P. M. 2005. "Shadow Analysis in High-Resolution Satellite Imagery of Urban Areas." *Photogrammetric Engineering and Remote Sensing* 71: 169–77.

Definiens eCognition. 2009. *Definiens eCognition Developer 8 Reference Book*. München: Definiens AG.

Dinis, J., A. Navarro, F. Soares, T. Santos, S. Freire, A. Fonseca, N. Afonso, and J. Tenedório. 2010. "Hierarchical Object-Based Classification of Dense Urban Areas by Integrating High Spatial Resolution Satellite Images and LiDAR Elevation Data." In *Proceeding of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (Vol. XXXVIII-4/C7), Geographic Object-Based Image Analysis (GEOBIA 2010)*, June 29–July 2, Ghent, Belgium.

Drăguţ, L., D. Tiede, and S. R. Levick. 2010. "ESP: A Tool to Estimate Scale Parameter for Multiresolution Image Segmentation of Remotely Sensed Data." *International Journal of Geographical Information Science* 24: 859–71.

Foody, G. M. 1999. "The Significance of Border Training Patterns in Classification by a Feedforward Neural Network Using Backpropagation Learning." *International Journal of Remote Sensing* 20: 3549–62.

Foody, G. M., and A. Mathur. 2004. "A Relative Evaluation of Multiclass Image Classification by Support Vector Machines." *IEEE Transactions on Geoscience and Remote Sensing* 42: 1335–43.

Foody, G. M., and A. Mathur. 2006. "The Use of Small Training Sets Containing Mixed Pixels for Accurate Hard Image Classification: Training on Mixed Spectral Responses for Classification by a SVM." *Remote Sensing of Environment* 103: 179–89.

Grigillo, D., and M. K. Fras. 2011. Classification Based Building Detection from GeoEye-1 Images. In *Joint Urban Remote Sensing Event (JURSE 2*011), edited by U. Stilla, P. Gamba, C. Juergens, and D. Maktav, 381–4. Munich: IEEE. doi:10.1109/JURSE.2011.5764799.

Haala, N., and C. Brenner. 1999. "Extraction of Buildings and Trees in Urban Environments." *ISPRS Journal of Photogrammetry and Remote Sensing* 54: 130–7.

Haralick, R. M., K. Shanmugam, and I. H. Dinstein. 1973. "Textural Features for Image Classification." *IEEE Transactions on Systems, Man, and Cybernetics* 3: 610–21.

Hermosilla, T., L. A. Ruiz, J. A. Recio, and J. Estornell. 2011. "Evaluation of Automatic Building Detection Approaches Combining High Resolution Images and LiDAR Data." *Remote Sensing* 3: 1188–210.

Hussain, E., S. Ural, K. Kim, C. Fu, and J. Shan. 2011. "Building Extraction and Rubble Mapping for City Port-Au-Prince Post-2010 Earthquake with GeoEye-1 Imagery and Lidar Data." *Photogrammetric Engineering and Remote Sensing* 77: 1011–23.

Im, J., J. R. Jensen, and J. A. Tullis. 2008. "Object-Based Change Detection Using Correlation Image Analysis and Image Segmentation." *International Journal of Remote Sensing* 29: 399–423.

Lang, S., F. Albrecht, S. Kienberger, and D. Tiede. 2010. "Object Validity for Operational Tasks in a Policy Context." *Journal of Spatial Science* 55: 9–22.

Liu, D., and F. Xia. 2010. "Assessing Object-Based Classification: Advantages and Limitations." *Remote Sensing Letters* 1: 187–94.

Longbotham, N., C. Chaapel, L. Bleiler, C. Padwick, W. J. Emery, and F. Pacifici. 2012. "Very High Resolution Multiangle Urban Classification Analysis." *IEEE Transactions on Geoscience and Remote Sensing* 50: 1155–70.

Lu, D., S. Hetrick, and E. Moran. 2010. "Land Cover Classification in a Complex Urban-Rural Landscape with QuickBird Imagery." *Photogrammetric Engineering and Remote Sensing* 76: 1159–68.

Marchisio, G., F. Pacifici, and C. Padwick. 2010. "On the Relative Predictive Value of the New Spectral Bands in the WorldView-2 Sensor." In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Honolulu, HI, July 25–30. Honolulu: IEEE. doi:10.1109/IGARSS.2010.5649771.

Marpu, P. R., M. Neubert, H. Herold, and I. Niemeyer. 2010. "Enhanced Evaluation of Image Segmentation Results." *Journal of Spatial Science* 55: 55–68.

Marshall, V., M. Lewis, and B. Ostendorf. 2012. "Do Additional Bands (Coastal, Nir-2, Red-Edge and Yellow) in WorldView-2 Multispectral Imagery Improve Discrimination of an Invasive Tussock, Buffel Grass (*Cenchrus ciliaris*)?" In *Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (Vol. XXXIX-B8), XXII ISPRS Congress*, August 25–September 1, Melbourne, Australia.

Mather, P. M. 2004. *Computer Processing of Remotely-Sensed Images*. 3rd ed. Chichester: Wiley.

Mathieu, R., J. Aryal, and A. K. Chong. 2007. "Object-Based Classification of Ikonos Imagery for Mapping Large-Scale Vegetation Communities in Urban Areas." *Sensors* 7: 2860–80.

Melgani, F., and L. Bruzzone. 2004. "Classification of Hyperspectral Remote Sensing Images with Support Vector Machines." *IEEE Transactions on Geoscience and Remote Sensing* 42: 1778–90.

Myint, S. W., P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng. 2011. "Per-Pixel vs. Object-Based Classification of Urban Land Cover Extraction Using High Spatial Resolution Imagery." *Remote Sensing of Environment* 115: 1145–61.

Piper, J. 1992. "Variability and Bias in Experimentally Measured Classifier Error Rates." *Pattern Recognition Letters* 13: 685–92.

Platt, R. V., and L. Rapoza. 2008. "An Evaluation of an Object-Oriented Paradigm for Land Use/Land Cover Classification." *The Professional Geographer* 60: 87–100.

Poli, D., E. Angiuli, and F. Remondino. 2010. "Radiometric and Geometric Analysis of WorldView-2 Stereo Scenes." In *Proceeding of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (Vol. XXXVIII-Part 1)*, June 15–18, Calgary, AB, Canada.

Pu, R., S. Landry, and Q. Yu. 2011. "Object-Based Urban Detailed Land Cover Classification with High Spatial Resolution IKONOS Imagery." *International Journal of Remote Sensing* 32: 3285–308.

Sawaya, K., L. Olmanson, G. Holden, J. Sieracki, N. Heinert, and M. Bauer. 2003. "Extending Satellite Remote Sensing to Local Scales: Land and Water Resource Management Using High Resolution Imagery." *Remote Sensing of Environment* 88: 143–55.

Shackelford, A. K., and C. H. Davis. 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: 2167–89.

Snedecor, G. W., and W. G. Cochran. 1980. *Statistical Methods*. 7th ed. Ames, IA: Iowa State University Press.

Song, M., D. Civco, and J. Hurd. 2005. "A Competitive Pixel-Object Approach for Land Cover Classification." *International Journal of Remote Sensing* 26: 4981–97.

Stumpf, A., and N. Kerle. 2011. "Object-Oriented Mapping of Landslides Using Random Forests." *Remote Sensing of Environment* 115: 2564–77.

Su, W., J. Li, Y. Chen, Z. Liu, J. Zhang, T. M. Low, S. Inbaraj, and A. M. H. Siti. 2008. "Textural and Local Spatial Statistics for the Object-Oriented Classification of Urban Areas Using High Resolution Imagery." *International Journal of Remote Sensing* 29: 3105–17.

Taşdemir, K., and S. Reis. 2011. "Land Cover Identification for Finding Hazelnut Fields Using WorldView-2 Imagery." In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Vancouver, July 24–29. Vancouver: IEEE. doi:10.1109/IGARSS.2011.6048922.

Tian, J., and D. M. Chen. 2007. "Optimization in Multi-Scale Segmentation of High-Resolution Satellite Images for Artificial Feature Recognition." *International Journal of Remote Sensing* 28: 4625–44.

Treitz, P., and P. J. Howarth. 2000. "High Spatial Resolution Remote Sensing Data for Forest Ecosystem Classification: An Examination of Spatial Scale." *Remote Sensing of Environment* 72: 268–89.

Tsai, F., and W. D. Philpot. 2002. "A Derivative-Aided Hyperspectral Image Analysis System for Land Cover Classification." *IEEE Transactions on Geoscience and Remote Sensing* 40: 416–25.

Turker, M., and K. San. 2010. "Building Detection from Pan-Sharpened Ikonos Imagery Through Support Vector Machines Classification." In *ISPRS Technical Commission VIII Symposium, Networking the World with Remote Sensing International, ISPRS Archives - Volume XXXVIII - Part 8*, edited by K. Kajiwara, K. Muramatsu, N. Soyama, T. Endo, A. Ono, and S. Akatsuka, 841–6. Kyoto: ISPRS.

Vinciková, H., M. Hais, J. Brom, J. Procházka, and E. Pecharová. 2010. "Use of Remote Sensing Methods in Studying Agricultural Landscapes – A Review." *Journal of Landscape Studies* 3: 53–63.

Yuan, F. 2008. "Land-Cover Change and Environmental Impact Analysis in the Greater Mankato Area of Minnesota Using Remote Sensing and GIS Modeling." *International Journal of Remote Sensing* 29: 1169–84.

Zerbe, L. M., and S. C. Liew. 2004. "Reevaluating the Traditional Maximum NDVI Compositing Methodology: The Normalized Difference Blue Index." In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium, Anchorage*, Alaska, September 20–24. Anchorage: IEEE. doi:10.1109/IGARSS.2004.1369774.

Zhou, X., T. Jancsó, C. Chen, and M. W. Verőné. 2012. "Urban Land Cover Mapping Based on Object Oriented Classification Using WorldView 2 Satellite Remote Sensing Images." Paper presented at the International Scientific Conference on Sustainable Development & Ecological Footprint, Sopron, March 26–27.