Influence of Data Source and Training Size on Impervious Surface Areas Classification Using VHR Satellite and Aerial Imagery Through an Object-Based Approach

Ismael Fernández, Fernando J. Aguilar, Manuel A. Aguilar, and M. Flor Álvarez

***Abstract*—Two very-high-resolution (VHR) satellite images from the GeoEye-1 and WorldView-2 sensors have been used in order to extract impervious surface areas (ISAs) over a Mediterranean coastal area of Almeria (Spain) through an object-based image analysis (OBIA). Different feature sets (basic multispectral information, relative spectral indices, and texture indices based on local variance) were used to feed a support vector machine (SVM) classifier in order to determine the most suitable information for ISAs classification. The classification results coming from both satellite images were compared to each other and also against those provided by a previous similar work carried out on an archival orthoimage. An estimation of the most appropriate number of training samples was performed for each data source by a sampling size reduction procedure. The accuracy assessment of the classification results showed that texture based on local variance was a valuable feature to improve ISA classification accuracy. When texture based on variance was included, the classification accuracy results provided by the archival orthoimage experiment (overall accuracy: 88.1% and KHAT: 0.760) were similar to those obtained from the VHR-satellite images (overall accuracy: 90.4% and 89.7%, KHAT: 0.806 and 0.792 for GeoEye-1 and WorldView-2, respectively). Finally, the influence of the data source and training size on ISA classification accuracy was also proved.**

***Index Terms*—Archival orthoimage, GeoEye-1, impervious surface area (ISA), object-based image analysis (OBIA), support vector machine (SVM), texture feature, very-high-resolution (VHR) satellite imagery, WorldView-2.**

I. Introduction

IMPERVIOUS surface areas (ISAs) are defined as anthropogenic features through which water cannot infiltrate into the soil [1]. ISAs lead to impacts on the environment by means of influencing hydrological and energy balances, affect the biological composition and functioning of ecosystems, and change the thermal properties of the soil [1], [2]. Therefore, the classification and monitoring of ISAs provide essential information needed in order to understand natural changes in ecosystems, especially in urban areas [3], [4]. Additionally, since ISAs constitute a good indicator of the degree of urbanization for a certain area, its classification in coastal areas is crucial to evaluate the effect of "coastal squeeze" (caused by artificially stabilized seafronts on sedimentary coastlines that inhibit the natural behavior of the coastal sedimentary processes) on coastal erosion processes. Because of their importance, efficient approaches for ISAs classification have been studied by the remote sensing scientific community.

According to [1], key issues to consider when remote sensing is used for ISAs classification are: spatial resolution, geometric characteristics of urban features, spectral resolution, and temporal resolution. Regarding the spatial resolution, a wide range of sensors have been tested, from medium or coarse spatial resolution satellite images, such as Landsat or MODIS [2], [5], [6], to very-high-resolution (VHR) images from satellites such as IKONOS [7] or Quickbird [8], and also including aerial orthoimages [9]. Second, and regarding the geometry of urban features, specific methods for road network extraction and building detection have been implemented, mainly because of the availability of VHR-satellite imagery and LiDAR data [10]. Third, ISAs have been extracted from a wide range of images with different spectral resolutions, ranging from three-band visible aerial orthoimages [9] to images from hyperspectral sensors [11]. Moreover, data from active sensors such as SAR image or LiDAR data have been also utilized to identify ISAs [12], [13]. It is worth noting that, from the knowledge of the authors, the influence of using different data sources for automatic ISAs classification for the same area has not been undertaken. Finally, regarding the temporal resolution and according to [1], the characteristics of the classes to be classified (vegetation, bare soil, buildings, road, etc.) the most suitable time of image acquisition can be determined.

The wide range of available data sources has led to different approaches for ISAs classification. Focusing on ISAs classification from VHR multispectral images, OBIA approaches have been widely used in order to minimize the "salt and pepper effect" by building meaningful segments or objects [14] and

Manuscript received January 20, 2014; revised April 05, 2014; accepted May 26, 2014. This work was supported in part by the Andalusia Regional Government, Spain, through the Excellence Research Project RNM-3575, in part by the Spanish Ministry for Science and Innovation (Spanish Government) under Grant Reference CTM2010-16573, and in part by the Cross-Border Cooperation Operational Programme Spain-External Borders POCTEFEX 2088-2013 under Grant Reference 0065_COPTRUST_3_E. All the three projects are cofinanced by the European Union through the European Regional Developed Funds (FEDER). This work takes part of the general research lines promoted by the CEI-MAR Campus of International Excellence as a joint initiative between the universities of Cadiz, Almeria, Granada, Huelva, and Malaga.

I. Fernández, F. J. Aguilar, and M. A. Aguilar are with the Department of Engineering, University of Almeria, Almeria, Spain (e-mail: ismaelf@ual.es; faguilar@ual.es; maguilar@ual.es).

M. F. Álvarez is with the Department of Mining Engineering, Area of Cartographic, Geodesy and Photogrammetric Engineering, University of Leon, Leon, Spain (e-mail: flor.alvarez@unileon.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2014.2327159

trying to improve the results of a more traditional per-pixel classification [15]. The OBIA approaches take advantage of the possibility of creating segments similar to real spatial objects such as buildings [1] and make the most of using contextual information and geometric attributes in the classification [14]. Regarding the classification stage, different algorithms have been utilized for the identification of ISAs when working with VHR images by means of OBIA approaches. Some examples are MANOVA-based algorithms [8], fuzzy rules based on spectral, spatial and texture attributes [7], minimum likelihood classifier (MLC) [16], classification and regression trees (CART) [17], [18], and data mining techniques [19]. In a previous research that the authors of this work conducted, three nonparametric classifiers were compared for ISAs classification on a VHR archival aerial orthoimage following an OBIA approach: CART, nearest neighbor (NN), and support vector machine (SVM), resulting with NN and SVM as the most accurate methods and the most similar to each other in terms of classification accuracy achieved (KHAT values 0.742 and 0.773 for NN and SVM while CART yielded 0.604) [9].

The aim of this work was to compare three VHR images, one archival aerial orthoimage (ArO) and two satellite images, one from GeoEye-1 and one from WorldView-2 (GE1 and WV2, respectively), for ISAs classification on a coastal area by means of an OBIA approach. Those data sources had different spatial resolutions (being the same for both satellites and finer for the ArO) and consisted of different spectral resolutions as well. Additionally, this work tried to identify the most suitable feature set for each data source regarding ISAs classification. Finally, the influence of the sampling size was also determined in order to find out how the sampling size reduction procedure affected each data source. This paper is organized as follows: first, the study area and the datasets used are presented; second, the methodology that was followed is explained; and third, the results are presented and discussed accordingly. Finally, the main conclusion is extracted and summarized.

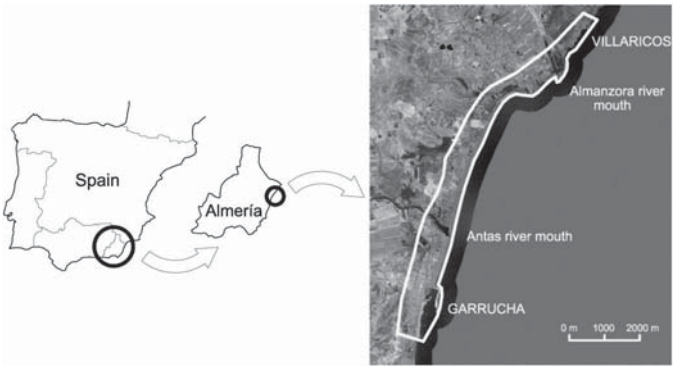


Fig. 1. Location of the study site on the Almeria coast, Southeastern Spain. F1:1

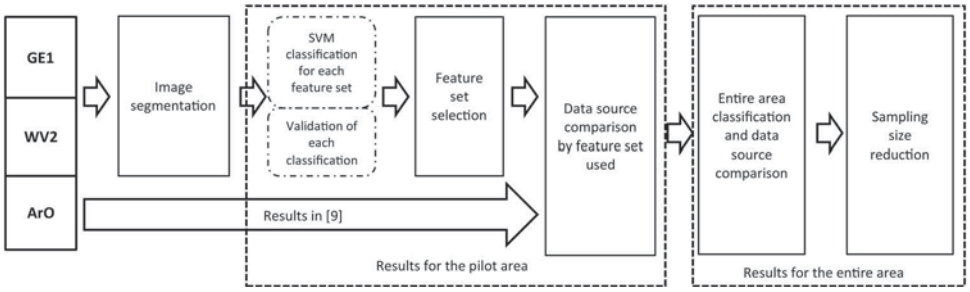
Fig. 2. Example of the urbanization process that occurred in the study area between 1977 (top image) and 2011 (bottom image). F2:1 F2:2

II. Study Area and Datasets

The study area consisted of a Mediterranean coastal area fringe that was approximately 11 000 m long and 775 m wide, corresponding to 841 ha and located in the province of Almeria (southeastern Spain, Fig. 1). This area has experienced a significant increase in ISA land cover since the 1970s, mainly because of a persistent sealing derived from a growth in urbanization, which is linked to a large development in the tourism industry of that area (Fig. 2), and it was chosen in order to determine the influence of the evolution of coastal hinterland ISAs land cover on the shoreline erosion that has occurred in this coastal area during the second half of the twentieth century.

For this work, GeoEye-1 and WorldView-2 images were used; both of them were acquired during August 2011 with only a 9-day gap between the images. Therefore, no significant land-cover changes were expected to occur between the images. Both of the images that were used in this work had a similar off-nadir angle (8.5° for GeoEye-1 and 10.0° for WorldView-2). Therefore, no differences in the results were expected due to that reason. Regarding spatial resolution, both satellite images provided a ground sample distance (GSD) of 0.5 m for the panchromatic band (PAN) and 2.0 m for the multispectral bands (MS). The main difference between both satellite images was the spectral resolution since GE1 only consisted of four bands (three bands in the visible spectrum plus one band in the near-infrared-NIR-) while WV2 provides four additional bands: coastal blue (CB), yellow (Y), red edge (RE), and a second NIR (NIR-2).

Two orthoimages were produced for each satellite image. First, one PAN orthoimage was derived from each image using the rational polynomial coefficients (RPCs), refined with seven ground control points (GCPs) acquired by a differential GPS (DGPS) field campaign (subdecimeter accuracy) by means of a simple translation at the image space [20], and a 1-m grid-spaced LiDAR-derived DEM ($\sigma_Z = \pm 0.09$ m estimated using 62 DGPS check points). The final RMSE_{2D} for the orthoimages that were produced, evaluated using 48 evenly distributed independent check points (ICPs), was 0.41 m for GeoEye-1 and



F3:1 Fig. 3. Workflow.

0.46 m for WorldView-2. This difference assured that there were no errors due to misallocation since it was smaller than the GSD. Second, one pan-sharpened orthoimage containing all the spectral information was derived from each satellite image (four or eight bands for GE1 or WV2, respectively). The pan-sharpened orthoimages had the same spatial resolution and accuracy as the PAN ones. For further and more detailed information about the orientation of the sensors, orthoimage generation, and quality assessment, refer to the previous works [20], [21].

III. Methods

The general workflow of this paper comprises the steps depicted in Fig. 3. First, an image segmentation procedure was carried out for each image. Then, a pilot area (the same as used in [9], located in the northern part of the images) was used in order to determine the most suitable feature set for each data source. For this purpose, a sampling design based on several subclasses for the target classes pervious and impervious was implemented, the feature sets to be tested were established, and an SVM classifier was used to identify the ISAs using each feature set. For this work, only the GE1 and WV2 feature set selections are presented since the results of the ArO image were stated in [9]. Then, all the three data sources were compared to each other regarding the feature set used. Once the most suitable feature set was determined for each image, the classification of the entire area was performed using the feature set selected and the data sources were compared again. Finally, a sampling size reduction was done for each image in order to test the suitable sampling size for each image.

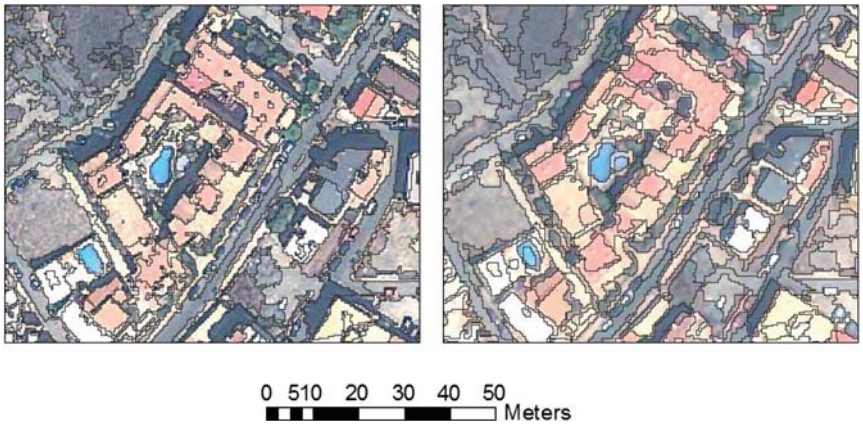
A. Segmentation Procedure

The widely used multiresolution segmentation algorithm implemented in eCognition 8 [22] was utilized to create the objects for the GE1 and WV2 images. This is a general segmentation algorithm based on homogeneity definitions in combination with local and global optimization techniques. A scale parameter is used to control the average image object size. It uses different homogeneity criteria for image objects based on spectral and/or spatial information. In this sense, the input parameters of this region-merging technique are scale, shape, and compactness and represent the size and internal heterogeneity of the objects. A detailed explanation about the multiresolution segmentation algorithm (eCognition approach) can be found in [22].

For the current work, all the MS bands of the pan-sharpened orthoimages were used and equally weighted for each satellite image. This means that four and eight bands were used for GE1 and WV2, respectively. Although this implies that both segmentations were different, the additional bands of the WV2 image were kept because they were later computed for the feature sets and, therefore, it was necessary to take into account those bands for the homogeneity parameters in the segmentation stage. The shape and compactness parameters were the same for the three datasets; they were 0.3 and 0.7, respectively. However, since GE1 and WV2 had the same spatial resolution, the scale parameter was chosen in order to yield similar segmentations from both images. Here, two criteria were taken into account: first, the object obtained should avoid under-segmentation errors (caused when objects cover more than one class [23]) and second, the size of the objects should be similar for both datasets since this ensures similar over-segmentation errors (then the classification error becomes independent of the segmentation and dependent on the classification strategy [23]). In summary, if the number of objects obtained from both image segmentations is similar, the size of objects could be considered similar too. Therefore, following these criteria, 50 and 40 were used as scale parameters by a "trial and error" procedure, by checking that the objects generally did not cover more than one class and verifying that the objects, which were created by the segmentation of the GE1 and WV2 images, were similar in size (Fig. 4). 120 410 and 116 788 objects were obtained for the GE1 and WV2 images, respectively (Fig. 4).

B. Sampling Design and Classification Strategy

Since the aim of this paper was to compare three VHR images regarding ISAs classification, it was necessary to identify two target classes in the imagery: 1) pervious and 2) impervious surfaces. However, they are not homogeneous classes and, therefore, a set of homogeneous subclasses were identified within each target class (8 and 10 subclasses for pervious and impervious classes, respectively, Table I), similar to those used in [9]. In order to follow the same strategy used in [9], the training and validation samples acquired for that experiment by means of a randomly stratified method were translated into the GE1 and WV2 images, updating their category when necessary but maintaining the same structure and spatial distribution. The classification of the ArO image was presented in [9]. Thus, a total amount of 583 training samples (320 pervious and 263 impervious) and 1783 validation samples (962 pervious and 821 impervious) were finally used for the GE1 and WV2 classifications for the entire area. Among these samples, 220 training samples and 728 validation samples were used for the pilot area with a distribution that is shown in Table I.



F4:1 Fig. 4. Example of the segmentation that was performed for GeoEye-1 (left image) and WorldView-2 (right image).

T1:1 TABLE I
TARGET CLASSES AND CORRESPONDING SUBCLASSES

Subclass	Target class	Training samples		Validation samples		Total training samples	Total validation samples
		GE1	WV2	GE1	WV2		
Sea	Pervious	46	46	76	76	146	462
Individual trees		16	16	28	28		
Bare soil		18	18	72	80		
Scrubland		20	20	112	103		
Beach		20	20	58	58		
Cultivated agricultural field		7	2	28	15		
Noncultivated agricultural field		9	14	40	54		
Forest		10	10	48	48		
Red building	Impervious	22	22	40	40	74	266
White building		7	7	31	31		
Gray building		4	4	14	14		
Road		13	13	51	51		
Path		7	7	24	24		
Harbor dam		2	2	22	22		
Sports court		1	1	5	5		
Swimming Pool		2	2	6	6		
Greenhouse		11	11	49	49		
Sidewalk		5	5	24	24		
Total sample size		220	220	728	728	220	728

Number of training and validation samples used for the classification and accuracy assessment of the pilot area for GE1 and WV2 classification. Note that the total number of training and validation samples was the same for both images.

Regarding the classification strategy and according to what was also found in [9], a direct classification of the target classes was used for the GE1 and WV2 image classifications since it resulted in a similar accuracy to that of the aggregation strategy, consisting of a classification of every subclasses and a subsequent aggregation of the results into the target classes. For direct classification of pervious and impervious classes, an SVM classifier was utilized since it was found to be the most efficient method in [9]. As a summary, the SVM was implemented by means of the free-distribution library LIBSVM [24] and following the methodology recommended by the authors, which consists of four steps: 1) a simple scaling of the training data; 2) a radial basis function (RBF) is used and the parameters C and γ are estimated by cross validation; 3) those parameters are used to test the dataset (by scaling the validation data in the same way that the training data was scaled) and the error matrix is computed (see [25] for more details); and 4) the computed SVM parameters are used to classify the scene.

C. Features Tested in Order to Classify Both Satellite Images

The feature sets explored for both satellite images, GE1 and WV2, were analogous to those used in [9] for the ArO image. Thus, the information was directly extracted from the spectral bands (basic spectral information); chromatic ratios (red, green, and blue ratios), green–red ratio used in the ArO experiment, and texture indices based on local variance were used in the GE1 and WV2 experiments. However, since NIR bands were available for both satellite images, NDIs were included for the GE1 and WV2 experiments. NDIs are one of the most common features used in remote sensing classification [e.g., normalized difference vegetation index (NDVI)] and they are based on the differences between one band from the visible spectrum and the NIR band

TABLE II
FEATURE SETS USED FOR GE1 AND WV2 EXPERIMENTS

Feature set name	GeoEye-1	WorldView-2
RGB*	R, G, B (3)	R, G, B (3)
Basic1	RGB set + PAN + Nir (5)	RGB set + PAN + Nir (5)
Basic2	-	Basic1 set + CB + Y + RE + Nir2 (9)
Rates1*	Basic1 set + Chromatic ratios (R, G, B) + G/R ratio (9)	Basic2 set + Chromatic ratios (R, G, B) + G/R ratio (13)
Rates2	Rates1 set + NDBI + NDGI + NDVI (12)	Rates1 set + NDBI + NDGI + NDVI + NDYI + NDREI + NDCBI (19)
Texture3	Rates2 + T3 (13)	Rates2 + T3 (20)
Texture5	Rates2 + T5 (13)	Rates2 + T5 (20)
Texture7	Rates2 + T7 (13)	Rates2 + T7 (20)
TextureAll*	Rates2 + T3 + T5 + T7 (15)	Rates2 + T3 + T5 + T7 (22)
TextureOnly	Basic1 set + T3 + T5 + T7 (8)	Basic2 set + T3 + T5 + T7 (12)

*Indicates that results are directly comparable with a similar feature set in the ArO experiment. The dimension of every feature set is indicated in brackets.

divided by the sum of both bands. It should be noted that only NIR-1 was used as the NIR reference information for WV2 since a high correlation between the NIR-1 and NIR-2 was detected ($R^2 = 0.995$). Thus, NDVI, NDGI, and NDBI (using red, green, and blue bands, respectively) were used for GE1 and WV2. Moreover, NDYI, NDREI, and NDCBI (using, respectively, the yellow, red edge, and coastal blue bands) were also computed for WV2.

Regarding the texture feature estimated from local variance, different window sizes were taken into account (3×3, 5×5, and 7×7). The size of the convoluted area was different from the one applied in the ArO experiment because of the different GSDs of the GE1 and WV2 PAN-orthoimages. In fact, while a 7×7 window size covered 1.4×1.4 m for the case of the ArO dataset, the same window size encompassed an area of 3.5×3.5 m for both GE1 and WV2 images, i.e., six times larger than the area corresponding to the same window size for the ArO image. According to this, the influence of the window size should be analyzed. Another difference between both experiments (ArO and VHR-satellite images) was the availability of the PAN band for the satellite images, which allowed the local variance to be computed from the PAN band. For the case of the ArO dataset, the local variance was estimated from each band of the visible spectrum (RGB). The resulting layers for GE1 and WV2 were called T3, T5, and T7, according to the window size used.

Summing up, four groups of feature sets were used (see Table II): 1) only RGB bands were used to be compared with the results of the ArO experiment; 2) basic information (Basic1 and Basic2, i.e., the RGB bands plus all the available bands for GE1 and WV2); 3) two types of rates: basic information plus the chromatic ratios (Rates1, comparable with a similar feature set in the ArO experiment), and rates which included NDIs (Rates2); and 4) several combinations of texture indices based on local variance: Rates2 plus T3, T5 or T7 (obtaining Texture3, Texture5, and Texture7), Rates2 plus all the different texture indices (obtaining TextureAll, comparable with a similar feature set in the ArO experiment called "Variance"), and Basic1 plus T3, T5, and T7 (in order to determine the influence of the Rates2 feature set).

D. Feature Set Selection

The most suitable feature sets for the GE1 and WV2 images were estimated by means of comparing the classification accuracy achieved by each feature set for the pilot area. Error matrices were calculated for every classification that was performed and the overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA), and KHAT statistic and its variance were derived [26]. Then, a comparison of each pair of classifications was performed by means of a Kappa analysis using the KHAT value and its variance for each classification [25]. For this work, significant statistical differences were considered at a 95% confidence level so two classification accuracies in terms of KHAT were considered as statistically different if the estimated Z statistic was larger than 1.96. Once the most appropriate feature set was selected, it was used to classify the entire area for each data source.

E. Data Source Comparison Regarding the Feature Set Used

In order to test the influence of the data source on the accuracy of the results, the pilot area classifications from the different VHR-images that used similar feature sets were compared to each other. These comparisons were also performed by means of a Kappa test. Thus, while 10 feature sets were used to compare GE1 and WV2 (Table III) only 3 feature sets could be utilized to compare ArO with both satellite images (RGB, Rates1, and TextureAll).

F. Entire Area Classification and Sampling Size Reduction

Once the most suitable feature set was selected for each dataset, that feature set was used to classify the entire area. Thus, classifications from the different data sources were again compared but using the entire sets of training and testing samples. Furthermore, a training size reduction (starting from the entire set of training samples) procedure was undertaken in order to test the optimum number of training samples for the ISAs classification for each data source. Nine different sizes of training sets were produced by randomly reducing the total samples in discrete steps of 10%, ranging from 10% to 90% of the total number of samples. Additionally, four repetitions for every

TABLE III
GENERAL ACCURACY RESULTS FOR THE GE1 STUDY

Feature set	OA	PA1	PA2	UA1	UA2	KHAT
RGB	86.1	95.9	69.2	84.4	90.6	0.685[a]
Basic1	87.8	95.9	73.7	86.4	91.2	0.725[ab]
Rates1	86.8	97.2	68.8	84.4	93.4	0.699[a]
Rates2	86.3	95.5	70.3	84.8	89.9	0.689[a]
TextureOnly	87.4	93.5	76.7	87.4	87.2	0.720[ab]
Texture3	88.6	96.5	74.8	86.9	92.6	0.744[ab]
Texture5	88.9	97.2	74.4	86.8	93.8	0.749[ab]
Texture7	87.5	94.6	75.2	86.9	88.9	0.721[ab]
TextureAll	90.0	93.7	83.5	90.8	88.4	0.781[b]

OA, PA, and UA values are expressed in %. 1 and 2 indicate the pervious and impervious classes, respectively. Along the KHAT column, figures presenting different letters indicate significant differences for the KHAT statistic ($p < 0.05$).

TABLE IV
GENERAL ACCURACY RESULTS FOR THE WV2 STUDY

Feature set	OA	PA1	PA2	UA1	UA2	KHAT
RGB	85.4	95.2	68.4	84.0	89.2	0.670[a]
Basic1	86.4	96.5	68.8	84.3	92.0	0.690[ab]
Basic2	87.4	95.5	73.3	86.1	90.3	0.716[ab]
Rates1	86.4	95.5	70.7	85.0	90.0	0.693[ab]
Rates2	88.0	89.6	85.3	91.4	82.5	0.744[abc]
TextureOnly	90.8	93.5	86.1	92.1	88.4	0.800[cd]
Texture3	89.0	94.2	80.1	89.1	88.8	0.758[bcd]
Texture5	90.8	96.3	81.2	89.9	92.7	0.796[cd]
Texture7	91.5	96.1	83.5	91.0	92.5	0.812[d]
TextureAll	90.8	95.0	83.5	90.9	90.6	0.798[cd]

OA, PA, and UA values are expressed in %. 1 and 2 indicate the pervious and impervious classes, respectively. Along the KHAT column, figures presenting different letters indicate significant differences for the KHAT statistic ($p < 0.05$).

sample size were performed by also using random methods. Therefore, the final number of different sample sets was 36 (4 repetitions for each of the 9 sample sizes). The number of samples corresponded with an approximate range of 0.05%–0.45% of the total number of objects created in the segmentation process. This decrease in the number of samples was performed only on the target classes pervious and impervious in order to simplify the process (not taking into account the subclasses). Thus, the correlation between the classification accuracy and the training size could be estimated for the ArO, GE1, and WV2 images. Moreover, the variability within each level of training size, referred to the statistical significance of the differences in accuracy within each size, was also assessed for each data source in order to determine the influence of the data distribution. Finally, the most suitable number of samples was estimated for each dataset taking into account that: first, the average accuracy achieved for the sampling size selected was not statistically different from that yielded by the total number of samples and second, no internal variability existed within the set of four repetitions of the selected size. With the first criterion, the maximum accuracy is assured (invariability regarding the sampling size), whereas with the second one the invariability regarding the sampling distribution is guaranteed.

IV. RESULTS AND DISCUSSION

A. GE1 and WV2 Feature Set Selection

The OA, PA, UA, and KHAT calculated from the error matrices for each feature set from GE1 and WV2 are shown in Tables III and IV, respectively. Those results highlighted that the feature sets yielded overall accuracies ranging from 86.1% to 90.0% for GE1 and from 85.4% to 90.8% for WV2. However, differences were found for both UAs and especially for the impervious PAs (up to a difference of 14.7% and 17.3% between the maximum and minimum classification accuracy figures for GE1 and WV2, respectively). On one hand, the results implied that while the PA for the pervious class was very high (around 95% for both images), the PA for the impervious class was significantly lower. The average of the difference between the PAs for the impervious and pervious classes was close to 22% and 16% for GE1 and WV2, respectively. On the other hand, the differences between the UAs for the two classes were much smaller for both images (the differences were about 5 percentage points each). According to these results, the pervious class was classified more accurately, especially if the omission error (related to PA) was considered.

Then, two kinds of confusions can be expected when the subclasses are considered: 1) intraclass confusion, which arises when two subclasses which correspond to the same target class (pervious or impervious) are misclassified, and 2) interclass confusion, which occurs when the misclassification happens between the two subclasses which belong to different target classes. Notice that only interclass confusion leads to error when the general error matrix is constructed since only two classes are considered: 1) pervious and 2) impervious (binary classification strategy). As an example, and taking into account an error matrix based on the subclasses for GE1 and WV2 and using the feature set TextureAll (results not shown), the interclass confusion was 30% and 24% of the total confusion for the pervious class in GE1 and WV2 classifications, respectively. On the other hand, the interclass confusion reached 52% and 54% of the total confusion for the impervious class for both satellite images, mainly because the impervious class was composed of a large number of spectrally different materials [16] and some of them were confused with some of the pervious subclasses. For instance, some impervious subclasses such as Gray building or Path were poorly classified (for WV2, PAs were 14% and 42%, and UAs were 22% and 44%, respectively). Since most of the confusion for both subclasses was bounded to pervious subclasses such as Beach and Bare Soil, these errors had to be added up for the general accuracy assessment of the final target classes pervious and impervious. Therefore, it was proved that impervious subclasses were misclassified as pervious ones more frequently than vice versa, leading to a low PA score for the impervious class (Tables III and IV).

The KHAT values highlighted that every feature set yielded a good agreement (> 0.60) and some of them were close to obtaining a strong agreement (> 0.80 according to [25]). The KHAT values were used to check for potential significant differences ($p < 0.05$) between the tested approaches through different Kappa tests. The results of this statistical analysis pointed out that for the GE1 experiment (Table III, last column on the right) only the TextureAll feature set showed statistically significant differences with respect to the other feature sets

(i.e., RGB, Rates1, and Rates2). However, the WV2 results yielded more statistical differences among the classifications and texture information improved the accuracy when compared to the feature sets that were based only on spectral information (both rates and basic spectral features), although no differences could be established among the feature sets based on texture (Table IV, last column on the right).

The results shown in Tables III and IV indicated that the KHAT might not be the most appropriate statistic to identify the best feature set. In fact, all the OA values seemed to be accurate and their range of variation was not too large. However, the difference between the PA and the UA for each class can be used as an indicator of the goodness of the classification accuracy results. Therefore, the smaller the difference, the more suitable the classification, since the PA and the UA would be more balanced. Regarding the GE1 results, whereas the Texture5 feature set yielded a similar OA to the TextureAll set, the latter yielded much smaller differences regarding the PA and the UA. Therefore, the TextureAll set was considered more suitable than the Texture5 set for GE1. On the other hand, Texture7 yielded the most accurate result and TextureOnly the most balanced one (taking into account PA and UA results) for WV2. However, TextureAll was also considered for WV2 in order to maintain the same feature set, but was only taken into account for comparison reasons for GE1 since its accuracy was not statistically different from Texture7 or TextureOnly feature sets.

For both the GE1 and WV2 classifications, it was found that the use of NDIs (Rates2 feature set) did not lead to a significant improvement in the accuracy if compared with simpler feature sets (i.e., fewer features). The obtained results were different from the results of a previous study in a less extensive urban zone in the same area, in which the NDIs ratios achieved more accurate results than the basic feature set [27] (OA from NDIs was 84.9% and 74.1% for GE1 and WV2, respectively, while OA from basic information was 81.5% and 72.1%, both being statistically significantly different). However, the classified area was not comparable since the classes that were targeted were significantly different. This contrasted with the fact that NDIs such as NDVI have been successfully utilized for the identification of ISAs, but mainly on medium and low spatial resolution images [28]. Furthermore, Tables III and IV showed that using different window sizes for the local texture variance calculation did not lead to statistically significant differences in the accuracy. Although a very large window size could lead to more significant differences, as it was found by [29] using IKONOS and Quickbird images when using up to a 21×21 window size, the window sizes that were implemented in this study were no larger than 7×7 since a rigorous analysis on the influence of window size was beyond the scope of this paper, and only the feature sets which could be compared with the ArO experiment were tested. Moreover, a combination of different window sizes, such as those used for the feature set TextureAll, turned out to have an improvement on the accuracy results (GE1 experiment). Hence, similar to [30]–[32], the local texture indices were the only feature that led to a significant increase in the accuracy and, therefore, the feature set TextureAll was applied to both VHR-satellite orthoimages in order to obtain the corresponding ISA classification.

TABLE V
SEPARABILITY OF THE RESULTS BETWEEN THE ARO EXPERIMENT AND BOTH THE GE1 AND WV2 EXPERIMENTS

Feature set	Data source	KHAT	Z statistic
RGB	ArO	0.507	
	GE1	0.685	**4.11**
	WV2	0.670	**3.73**
Rates1	ArO	0.554	
	GE1	0.699	**3.45**
	WV2	0.693	**3.23**
Texture	ArO	0.776	
	GE1	0.781	0.16
	WV2	0.798	0.67

Values above 1.96 (bold figures) indicate significant differences for the KHAT statistic between ArO and each satellite image experiment.

B. VHR-Satellite Images Versus Archival Orthoimage for Comparable Feature Sets

The results of the comparison are depicted in Table V. They showed that ArO yielded a less accurate classification than any of the satellite datasets when using the RGB and Rates1 feature sets. The results also proved that the implementation of local variance texture indices improved the classification regardless of the considered imagery, but it was quantitatively larger in the case of ArO. This results agrees with the fact that the improvement of classification accuracy by means of applying texture indices depends on the spatial resolution [30], being more effective for finer resolution images [30], [33]. Furthermore, it was demonstrated that the classifications which used feature sets with local variance textures yielded similar accuracy results (i.e., they were not statistically different).

On one hand, it is important to highlight that when the texture local variance features were considered, the data with higher radiometric quality (i.e., GE1 and WV2) did not lead to a significant improvement in the classification compared to using higher spatial resolution archival RGB orthoimage data (which included some artefacts and poor radiometry). In fact, in other studies such as [17], high-resolution aerial orthoimage data achieved more accurate results than Quickbird VHR data, although a different classifier and spatial resolution were used. On the other hand, it was also proved that both satellite images yielded significantly higher accuracy when only basic information and ratios between bands were used. This could mean that these kinds of satellite images may be radiometrically more suitable for image classification due to the difficulty of accurately classifying a heterogeneous image such as the image that was used in ArO experiment (spatial resolution of 0.20 m).

As a result, these findings underlined the importance of using invariant features for image classification, such as the local variance texture, especially when VHR images are used [33]. Another work carried out using an SVM classifier and an OBIA approach on a Quickbird image by optimizing the scale parameter obtained a maximum OA of about 90% [23]; this was similar to our results.

C. GeoEye-1 Versus WorldView-2 Image Classification

In order to check the influence of the satellite imagery source on the extraction of ISAs, the classification accuracy results

TABLE VI
T6:1 SEPARABILITY OF THE RESULTS BETWEEN THE GE1 AND WV2 EXPERIMENTS

Feature set	Data source	KHAT	Z statistic
RGB	GE1	0.685	0.38
	WV2	0.670	
Basic1	GE1	0.725	0.90
	WV2	0.690	
Basic2	WV2	0.716	0.24
Rates1	GE1	0.699	0.15
	WV2	0.693	
Rates2	GE1	0.689	1.43
	WV2	0.744	
Texture3	GE1	0.744	0.39
	WV2	0.758	
Texture5	GE1	0.749	1.35
	WV2	0.796	
Texture7	GE1	0.721	**2.58**
	WV2	0.812	
TextureOnly	GE1	0.720	**2.25**
	WV2	0.800	
TextureAll	GE1	0.781	0.51
	WV2	0.798	

Note that Basic1 and Basic2 feature sets for WV2 were compared with Basic1 set for GE1. Values above 1.96 (bold figures) indicate significant differences for the KHAT statistic between both accuracy figures ($p < 0.05$).

achieved using GE1 and WV2 were compared over the pilot area (Table VI). Those results showed that the differences between both image sources were small. Actually, they were only statistically significant when the feature sets Texture7 and TextureOnly were considered. For these cases, WV2 showed a higher accuracy. Although the differences were not significant, WV2 yielded higher values in the classification accuracy when the additional bands (Yellow, Coastal Blue, Red Edge, and Nir-2) and the ratios derived from those bands (Rates2) were used. When not considering these additional bands, GE1 obtained slightly more accurate classification results than WV2. It should be noted that the difference between using those additional bands or not (Basic1 and Basic2 feature sets) was not statistically significant for WV2, so the effect of including those bands was only perceptible when the texture indices (Texture7 and TextureOnly) were included.

It is worth noting that these results are slightly different from those achieved in a previous study carried out on a smaller urban area of the same study site [27] in which GeoEye-1 imagery led to more accurate results than WorldView-2 imagery. Although not exactly the same feature set was used (e.g., the local variance texture was not considered in the previous study), and also a different classifier was tested, it seems to be clear that the classes that are targeted and the use of a larger and more heterogeneous area may affect the results of the satellite image comparison.

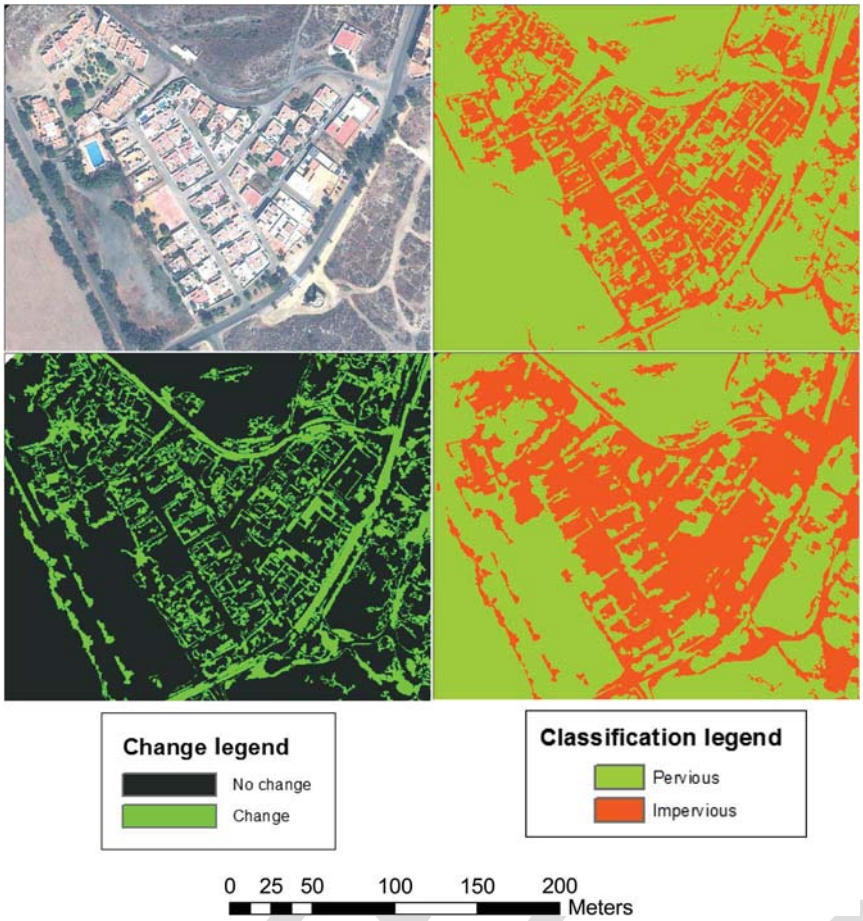
Some differences can be observed when the image classifications from both satellites using TextureAll are compared for the pilot area (Fig. 5). First, it should be stated that 12.8% of the image pixels were classified differently in both images if only the target classes are considered. Many of them occurred on the edges of buildings due to the differences in shadows since the solar elevation were not the same for both satellite images, that being shadows were larger for the GE1 image. Additionally, since the top of the buildings had different orthorectification errors for both satellites, these edge errors increased. However, some differences were not related to the image features, but to the classification method. Thus, more ISAs were estimated by means of WV2 (13.9% from the whole pilot area of ISAs for GE1 and 20.7% for WV2). For instance, it can be seen from Fig. 5 how GE1 misclassified a part of a road (PA was 67% for GE1 while it was 88% for WV2). Moreover, it was proved that the interclass confusion among subclasses was highly variable depending on the satellite image. On one hand, among pervious subclasses, while Individual Trees and Cultivated agricultural field were more confusing for GE1, Beach was less accurately classified by WV2. On the other hand, the impervious subclasses Road, Harbour dam, and Side walk yielded more interclass confusion for GE1, whereas Gray Building and Path led to more confusion for WV2. This fact meant that although the overall accuracies (taking into account the target classes pervious and impervious) were similar for both satellite images, the distribution of the classification errors was different; this led to different classification results as well.

D. Estimation of the Most Efficient Training Set

First, the classification accuracy results from each source obtained using all the training and testing samples from the whole working area (instead of from only the pilot area) and applying the TextureAll feature set are discussed (Table VII). In this sense, all the tested data sources achieved high accuracies, presenting an OA close to 90%. Furthermore, the PA and UA figures were balanced for both the pervious and impervious classes (i.e., the small differences between them). These results proved that a large and well-distributed training set could lead to an accurate classification regardless of the type of image used. In fact, only the GE1 classification results were statistically more accurate than the ArO results, whereas the WV2 ones were not significantly different from any of the others. However, since a relatively large number of training samples were used, and taking into account that this is not usually affordable in an operational workflow, it was thought to be important to test how the training sample size reduction would affect the classification accuracy obtained from each data source. Actually, it has been demonstrated that the number of training samples can be crucial for classification accuracy estimations [34]. Additionally, nonparametric classifiers such as SVM may not require a large amount of training samples but they need appropriate samples, i.e., those lying on the edge of the class distribution or support vectors [35].

As an example, the results of the training sample size reduction procedure for the GE1 experiment are depicted in Fig. 6. This figure showed that the OA tended to increase when more samples were used. Likewise, the variability of the OA values was reduced when larger training sample sizes were employed. Similar findings could be found for the ArO and WV2 experiments (figures not shown).

In the case of the ArO experiment, although the average OA for the 50% training level was not statistically different from the classification that was based on all the samples, the variability within repetitions was not reduced until 60% of the total training



F5:1 Fig. 5. Comparison between GE1 and WV2 classifications using TextureAll feature set over a sample zone of the pilot area. Top-left: original GE1 image. Top-right:
F5:2 GE1 classification. Bottom-right: WV2 classification. Bottom-left: GE1-WV2 differences.

TABLE VII
T7:1 GENERAL ACCURACY RESULTS FOR ARO, GE1, AND WV2 EXPERIMENTS WHEN ALL
T7:2 TRAINING AND TESTING SAMPLES WERE INCLUDED AND THE TEXTUREALL FEATURE SET
T7:3 WAS APPLIED

	Archival orthoimage	GeoEye-1	WorldView-2
OA	88.1	90.4	89.7
PA1	88.6	90.5	90.9
PA2	87.5	90.1	88.3
UA1	89.2	91.5	90.1
UA2	86.8	89.0	89.2
KHAT	0.760^{b}	0.806^{a}	0.792^{ab}
Training size (no. objects)	576	583	583
Testing size (no. objects)	1796	1783	1783

OA, PA, and UA are expressed in %. Significant differences ($p < 0.05$) between KHAT results are indicated by different superscript letters.

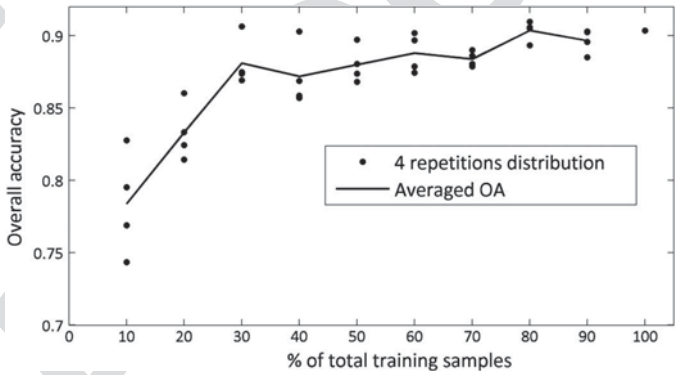


Fig. 6. OA for each repetition and the average value for each training sample size. GE1 experiment. F6:1 F6:2

samples were used. Therefore, the latter percentage was chosen as the appropriate number of training samples, being 346 the minimum number of samples to use from a total of 576 samples. Regarding the GE1 dataset, the 60% training level ensured a similar accuracy to the total sample set, but 70% of the samples were needed to be used in order to obtain invariability among repetitions. Thus, using 408 out of 583 samples was considered to be suitable for GE1, which meant it had more than 60 additional samples compared to the ArO trial. Finally, for the WV2 dataset, when 60% of the total of training samples were used, it did not show any variability among the four repetitions, although the average OA turned out not to be significantly different once the 40% level was reached. Therefore, 60% of the total training samples were needed to ensure the maximum accuracy. This meant that 350 out of 583 samples had to be used.

In addition to the previous analysis, the performance of the data source was compared by estimating the statistical

TABLE VIII
AVERAGE RESULTS REGARDING OVERALL CLASSIFICATION ACCURACY FOR EVERY DATA SOURCE AND PERCENTAGE OF TOTAL TRAINING SAMPLES LEVEL (EXPRESSED IN %)

% of total samples	ArO	GE1	WV2
10	82.28[a]	**78.39[b]**	82.67[a]
20	**82.29[b]**	**83.31[b]**	85.78[a]
30	**84.59[b]**	88.11[a]	87.39[a]
40	**85.38[b]**	87.18[a]	88.46[a]
50	87.17	88.00	88.11
60	87.37	88.80	89.29
70	86.65	88.39	88.73
80	**87.61[b]**	90.35[a]	89.01[ab]
90	**87.15[b]**	89.65[a]	89.55[a]
100	**88.08[b]**	90.35[a]	89.68[ab]

Significant differences ($p < 0.05$) between OA results situated along rows (data sources) are indicated by different superscript letters. Bold figures indicate the less accurate and statistically different result ($p < 0.05$) for every training sampling size.

separability ($p < 0.05$) of their average OAs with the differences between every two proportions [36]. The OAs were used instead of the Kappa since an average Kappa from the four repetitions would have been required. The results can be reviewed in Table VIII. The ArO classifications were less accurate than some of the classifications using satellite imagery, either when a large percentage of samples were utilized (80%–100%) or when few samples were used (20%–40% of the total training samples). An exception occurred when 10% of the total samples were used, where the ArO experiment yielded similar accuracy results to WV2, and both being more accurate than GE1. Otherwise, no differences were found among any data source when using the intermediate percentages (50%–70%). The results for both satellite image experiments were only statistically different for the lowest percentages (10% and 20%), where WV2 led to more accurate results than GE1. Those results showed that the use of scanned archival aerial orthoimages generally tends to require a larger sample size in order to obtain accuracies similar to the ones from satellite images.

According to the aforementioned results, the SVM algorithm has been proved as a suitable approach for classifying impervious areas for all of the image sources that were tested. An inherent relationship was also found between the accuracy results and the training sample size for the SVM approach; this is similar to the findings of [37]. In fact, the accuracy achieved with a certain training sample may lead to biased results since it has been proved that not only the size but also the distribution of the sample matters.

Regarding the sample size reduction, it should be noted that it was performed without taking into account any stratifications by subclasses and that only the classes, pervious and impervious, were considered. Hence, the probability of removing key samples was high when low percentages of samples were used and, therefore, the final accuracy could have probably been negatively affected. Furthermore, since a nonclassifier-oriented training selection was performed (as suggested by [34], [38]), the more the samples removed, the higher the likelihood of eliminating meaningful support vectors that would be needed to effectively separate pervious and impervious classes.

V. CONCLUSION

A classification strategy based on one approach developed previously for an archival orthoimage was found to be suitable when using it on GeoEye-1 and WorldView-2 VHR-satellite imagery for the classification of ISAs. That was an important finding since it enables a similar approach over different types of data sources which facilitate multitemporal and multisource land-cover studies to be performed. Both satellite images led to higher classification accuracies than the archival RGB orthoimage when texture was not included. However, when local variance texture was added, the accuracy results for all the data sources were not statistically different. Slight differences were found between both satellite image classifications for each one of the feature sets tested, so it cannot be concluded that the higher number of bands available in the WorldView-2 images provided a significantly more accurate classification than the one obtained from GeoEye-1.

When the most numerous and well-distributed training samples were used to train the SVM classifier on the entire study area, the GeoEye-1 classification yielded significantly higher accuracies than those obtained from the archival orthoimage. The classification accuracy results that corresponded to the WorldView-2 satellite image were not statistically different with respect to any other data source. Moreover, the experiment related to the reduction of training samples clearly demonstrated that the training size and its distribution over the working area played a key role. In this sense, it was proved that larger uncertainties in the classification results could be obtained if the number of samples is not appropriate.

Finally, SVM was found to be a suitable classifier that can separate pervious and impervious classes when a proper training set is available and when invariant feature sets such as texture based on local variance are used.

REFERENCES

[1] Q. Weng, "Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends," *Remote Sens. Environ.*, vol. 117, no. 15, pp. 34–49, Feb. 2012.
[2] J. O. Sexton *et al.*, "Urban growth of the Washington, D.C.-Baltimore, MD metropolitan region from 1984 to 2010 by annual, Landsat-based estimates of impervious cover," *Remote Sens. Environ.*, vol. 129, pp. 42–53, Feb. 2013.
[3] H. Ding and W. Shi, "Land-use/land-cover change and its influence on surface temperature: A case study in Beijing City," *Int. J. Remote Sens.*, vol. 34, no. 15, pp. 5503–5517, Aug. 2013.
[4] X. Zhang, D. Pan, J. Chen, Y. Zhan, and Z. Mao, "Using long time series of Landsat data to monitor impervious surface dynamics: A case study in the Zhoushan Islands," *J. Appl. Remote Sens.*, vol. 7, no. 1, pp. 073515-1–073515-14, Jan. 2013.
[5] T. E. Parece and J. B. Campbell, "Comparing urban impervious surface identification using landsat and high resolution aerial photography," *Remote Sens.*, vol. 5, no. 10, pp. 4942–4960, Oct. 2013.
[6] C. Deng and C. Wu, "The use of single-date MODIS imagery for estimating large-scale urban impervious surface fraction with spectral mixture analysis and machine learning techniques," *ISPRS J. Photogramm. Remote Sens.*, vol. 86, no. 1, pp. 100–110, Dec. 12, 2013.
[7] X. Hu and Q. Weng, "Impervious surface area extraction from IKONOS imagery using an object-based fuzzy method," *Geocarto Int.*, vol. 26, no. 1, pp. 3–20, Feb. 2011.

[8] Y. Zhou and Y. Q. Wang, "Extraction of impervious surface areas from high spatial resolution imagery by multiple agent segmentation and classification," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 7, pp. 857–868, Jul. 2008.
[9] I. Fernandez Luque, F. J. Aguilar, M. F. Alvarez, and M. A. Aguilar, "Non-parametric object-based approaches to carry out ISA classification from archival aerial orthoimages," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 4, pp. 2058–2071, Aug. 2013.
[10] M. E. Hodgson, J. R. Jensen, J. A. Tullis, K. D. Riordan, and C. M. Archer, "Synergistic use of lidar and color aerial photography for mapping urban parcel imperviousness," *Photogramm. Eng. Remote Sens.*, vol. 69, no. 9, pp. 973–980, Sep. 2003.
[11] Q. Weng, X. Hu, and D. Lu, "Extracting impervious surfaces from medium spatial resolution multispectral and hyperspectral imagery: A comparison," *Int. J. Remote Sens.*, vol. 29, no. 11, pp. 3209–3232, Jun. 2008.
[12] Y. Zhang, H. Zhang, and H. Lin, "Improving the impervious surface estimation with combined use of optical and SAR remote sensing images," *Remote Sens. Environ.*, vol. 141, pp. 155–167, Feb. 2014.
[13] X. Zhu and T. Toutin, "Land cover classification using airborne LiDAR products in Beauport, Québec, Canada," *Int. J. Image Data Fusion*, vol. 4, no. 3, pp. 252–271, Jul. 2013.
[14] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 2–16, Jan. 2010.
[15] S. W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng, "Per-pixel vs. object-based classification or urban land cover extraction using high spatial resolution imagery," *Remote Sens. Environ.*, vol. 115, no. 5, pp. 1145–1161, May 2011.
[16] D. Lu, S. Hetrick, and E. Moran, "Impervious surface mapping with QuickBird imagery," *Int. J. Remote Sens.*, vol. 32, no. 9, pp. 2519–2533, May 2011.
[17] L. Matikainen and K. Karila, "Segment-based land cover mapping of a suburban area-comparison of high-resolution remotely sensed datasets using classification trees and test field points," *Remote Sens.*, vol. 3, no. 8, pp. 1777–1804, Aug. 2011.
[18] R. Pu and S. Landry, "A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species," *Remote Sens. Environ.*, vol. 124, no. 1, pp. 516–533, Sep. 2011.
[19] C. M. D. de Pinho, L. M. G. Fonseca, T. S. Korting, C. M. de Almeida, and H. J. H. Kux, "Land-cover classification of an intra-urban environment using high-resolution images and object-based image analysis," *Int. J. Remote Sens.*, vol. 33, no. 10, pp. 5973–5995, Oct. 2012.
[20] M. A. Aguilar, M. D. M. Saldaña, and F. J. Aguilar, "Assessing geometric accuracy of the orthorectification process from GeoEye-1 and WorldView-2 panchromatic images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 21, pp. 427–435, Jun. 2013.
[21] M. A. Aguilar, F. J. Aguilar, M. M. Saldaña, and I. Fernández, "Geopositioning accuracy assessment of GeoEye-1 panchromatic and multispectral imagery," *Photogramm. Eng. Remote Sens.*, vol. 78, no. 3, pp. 247–257, Mar. 2012.
[22] U. C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen, "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 3–4, pp. 239–258, Jan. 2004.
[23] D. Liu and F. Xia, "Assessing object-based classification: Advantages and limitations," *Remote Sens. Lett.*, vol. 1, no. 4, pp. 187–194, Dec. 2010.
[24] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.
[25] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data*, 2th ed. Boca Raton, FL, USA: CRC Press, 2009.
[26] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.*, vol. 37, no. 1, pp. 35–46, Jul. 1991.
[27] M. A. Aguilar, M. M. Saldaña, and F. J. Aguilar, "GeoEye-1 and WorldView-2 pan-sharpened imagery for object-based classification in urban environments," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2583–2606, Apr. 2013.
[28] D. Lu, H. Tian, G. Zhou, and H. Ge, "Regional mapping of human settlements in southeastern China with multisensor remotely sensed data," *Remote Sens. Environ.*, vol. 112, no. 8, pp. 3668–3679, Sep. 2008.
[29] L. Wang, W. P. Sousa, P. Gong, and G. S. Biging, "Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama," *Remote Sens. Environ.*, vol. 91, no. 3-4, pp. 432–440, Jun. 2004.
[30] D. Chen, D. A. Stow, and P. Gong, "Examining the effect of spatial resolution and texture window size on classification accuracy: An urban environment case," *Int. J. Remote Sens.*, vol. 25, no. 11, pp. 2177–2192, Jun. 2004.
[31] A. S. Laliberte and A. Rango, "Texture and scale in object-based analysis of subdecimeter resolution unmanned aerial vehicle (UAV) imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 1–10, Mar. 2009.
[32] T. Ota, N. Mizoue, and S. Yoshida, "Influence of using texture information in remote sensed data on the accuracy of forest type classification at different levels of spatial resolution," *J. Forest Res.*, vol. 16, no. 6, pp. 432–437, Dec. 2011.
[33] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, Jan. 2007.
[34] G. M. Foody and A. Mathur, "Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification," *Remote Sens. Environ.*, vol. 93, no. 1-2, pp. 107–117, Oct. 2004.
[35] G. M. Foody, A. Mathur, C. Sanchez-Hernandez, and D. S. Boyd, "Training set size requirements for the classification of a specific class," *Remote Sens. Environ.*, vol. 104, no. 1, pp. 1–14, Sep. 2006.
[36] G. M. Foody, "Sample size determination for image classification accuracy assessment and comparison," *Int. J. Remote Sens.*, vol. 30, no. 20, pp. 5273–5291, Oct. 2009.
[37] G. M. Foody and A. Mathur, "A relative evaluation of multiclass image classification by support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1335–1343, Jun. 2004.
[38] A. Mathur and G. M. Foody, "Crop classification by support vector machine with intelligently selected training data for an operational application," *Int. J. Remote Sens.*, vol. 29, no. 8, pp. 2227–2240, Apr. 2008.

Ismael Fernández received the Ph.D. degree from University of Almeria, Almeria, Spain, in 2013.

He is currently working as a Postdoctorate Researcher. He is also a Peer-Reviewer of remote sensing and coastal management journals. His research interests include remote sensing classification and impervious areas extraction.

Fernando J. Aguilar received the Ph.D. degree from University of Cordoba, Cordoba, Spain, in 1997.

He is currently a Full Professor with the Department of Engineering, University of Almeria, Almeria, Spain. At present, he is the Head of the Research Group called "Integrated Landscape Management and Spatial Information Technologies," leading and collaborating in research projects related to geospatial data fusion, monitoring natural resources and mapping from VHR satellite data. He is the coauthor of several international publications, and is also a Peer-Reviewer of many of the main GIS and remote sensing journals all over the world. His research interests include digital elevation models, airborne laser scanning, digital photogrammetry, and GIS.

Manuel A. Aguilar received the Agricultural Engineering and the Ph.D. degrees from the University of Córdoba, Córdoba, Spain, in 1996 and 2001, respectively.

Since 1999, he has been a Professor of Technical Drawing with the Almería Engineering High School, University of Almería, Almería, Spain. He is currently a Senior Lecturer with the Department of Engineering, University of Almería. His research interests include close-range, satellite and classic digital photogrammetry, remote sensing of land cover, as well as the generation and quality control of DEMs.

M. Flor Álvarez received the Ph.D. degree from the University of Vigo, Vigo, Spain, in 2006.

She currently works as an Assistant Professor with the Department of Mining Engineering, University of León, León, Spain. She has been a Guest Researcher at the Pacific Forestry Centre, BC, Canada and at the ITC, Eanske, The Netherlands. She is the coauthor of several international publications and is a Peer-Reviewer of several GIS and remote sensing journals. Her research interests include monitoring natural resources using remote sensing.