An adaptive approach for simultaneous classification of remote sensing scenes including rural and urban targets

Leticia Sartorio \(^b\) and Daniel Zanotta \(^a\)

\(^a\)Center for human and information sciences, University of Rio Grande, Rio Grande, Brazil; \(^b\)Advanced Visualization & Geoinformatics Laboratory, UNISINOS University, São Leopoldo, Brazil

**ABSTRACT**

In this paper, an automatic adaptive image classification framework designed to operate in multiresolution scenes including rural and urban targets is proposed and tested. Traditional image analysis is commonly aimed to classify images using a single strategy and source of data over the entire scene. Ideally, urban targets should predict specialized classification systems using high spatial resolution images, such as object-based image analysis and non-parametric classifiers. Conversely, rural targets should be handled with the high-spectral resolution, pixel-based classification approaches, and parametric techniques. The formulation proposed in this study starts by performing an prior separation of rural and urban areas by assuming Central Limit Theorem (CLT) establishments. Then, both kinds of targets are labelled in an automatic adaptive fashion, each one with proper data and method previously selected. One experiment performed using set of data composed of a high spatial resolution true-colour image and a multispectral image, as well as preselected classification techniques particularly adjusted for each case. Visual and quantitative assessing by two accuracy metrics testing the proposed approach versus traditional classification confirm the soundness of the proposed framework.

**KEYWORDS**

Classification; remote sensing; image segmentation; pattern recognition; landscape

**ARTICLE HISTORY**

Received 2 October 2019
Accepted 12 December 2019

**CONTACT**

Daniel Zanotta \(\text{dzanotta@edu.unisinos.br}\)
UNISINOS University, São Leopoldo, Brazil

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group on behalf of the International Water, Air & Soil Conservation Society(INWASCON). This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Introduction**

The increasing development of sophisticated remote sensing instruments has conferred considerable improvements in the quality of images acquired from space (Gholoobi & Kumar, 2015). Public administration has been facing a growing dependence of rapid and reliable monitoring of an increasingly dynamic and complex scenario. High spatial resolution imaging sensors onboard satellites are one of the examples which have allowed detailed land use and land cover mapping at a high efficiency and relatively low cost (Fisher, Eileen, James Dennedy-Frank, Kroeger, & Boucher, 2017), mainly over urban areas and other complex environments. Missions that brought advances in this direction are, in chronological order, IKONOS, QuickBird, RapidEye, Geoeye, WorldView (Chuvieco, 2016), as well as aircraft and the recent unmanned aerial vehicle (UAV) images. The presence of complex ground targets is common in this type of high spatial resolution images but can be adequately classified by modern computational techniques like Support Vector Machines (SVM), Random Forests (RF), and, more recently, by Convolutional Neural Networks (CNN) (Jensen, 2009). Indeed, these are the most common image classification techniques recently found in the literature.

Automatic image classification can be performed at the pixel level (pixel-based), where each pixel is analyzed individually for labelling, or at the object level (object-based), where a set of pixels is previously merged receiving a single label (Moosavi, Talebi, & Shirmohammadi, 2014). The pixel-based classification is restricted to use only the spectral information of pixels as the unique attribute, not considering any other aspect in the process (Weih & Riggan, 2010). For many applications, this approach is able to retrieve a thematic map showing the elements of interest throughout the image with reasonable precision. However, in more complex applications, e.g. involving small structures or well elaborated shapes, like urban areas or data including detailed targets like some agricultural areas or lithological mapping, the object-based classification is more suitable (Zhou, Troy, & Grove, 2008). The reason is that object-based approach is performed in two basic steps: image segmentation, that aims to group similar pixels in objects, and classification, that aims to label the resulting objects (Whiteside, Boggis, & Maier, 2011). Working with objects allows the analyst to explore not only radiometric information, as in pixel-based approach, but also attributes like texture, shape, size, and context, improving the classification process (Duro, Franklin, & Dublé, 2012). Indeed, the object-based approach is able to take advantage of surrounding and circumstantial characteristics like roughness, neighborhood, size, and morphology of resulting objects.
Due to the above-mentioned reasons, the latest consensus of the specialized literature indicates that urban targets (buildings, roads, trees, small waterbodies) present in high spatial resolution images should be classified by objects (Bhaskaran, Paramananda, & Ramnarayan, 2010; Ma et al., 2017), since detailed information about shape, texture, and context are very important attributes to differentiate among targets these standard attributes (Mather & Tso, 2016). Object-based approach along with hierarchical and nonparametric classifier (no assumption of probability distribution of classes), is expected to be more effective in the urban scenes. Conversely, rural targets (fields, forests, medium and large waterbodies, rocks, and varied soils) ought to be better classified at pixel-level, with images including as many as spectral bands are possible (Aguirre-Gutiérrez, Seijmonsbergen, & Duivenvoorden, 2012; Ferreira, Zortea, Zanotta, Shimabukuro, & Souza Filho, 2016). Rural or natural targets present large areas and subtle radiometric variations along distances, which allows correct description by even very low spatial resolution images. Due to this reason, sensors designed to monitor these areas usually have many spectral bands (Fisher et al., 2017), enabling detailed description about the chemical composition of targets, which is crucial for lithological or vegetation mapping (Herold, Roberts, Gardner, & Dennison, 2004; Lillesand, Kiefer, & Chipman, 2014). Furthermore, for rural targets, parametric classifiers (which assume well-known probability distribution of classes) might estimate the classes with greater efficiency, since it is designed to work with only radiometric information, which is very abundant in multispectral images of medium to low spatial resolution (Whiteside et al., 2011).

There are several strategies for improving and refine the classification of high spatial resolution and hyperspectral images (Zhao, Du, & Emery, 2017; Zhong et al., 2017; Zhong, Ma, Ong, Zhu, & Zhang, 2017). Despite the recent improvements, the classification of scenes simultaneous including urban and rural targets remains a challenge for classifiers traditionally used for remote sensing image recognition. As can be noted this condition is very common and brings many challenges when dealing with mapping heterogeneous areas. Analysts usually rely on the time-consuming and labor-intensive prior separation of the different targets by visual interpretation followed by independent classification of each area. An alternative is the selection of a unique method retrieving the best trade-off using the high spatial resolution image available. However, the use of only one type of image data and classification strategy in these mixed scenarios hinders the optimization of the accuracy of the results.

The present study suggests an automatic adaptive framework to overcome the classification of scenes simultaneously including rural and urban classes. It assumes the use of two different images covering the same area: one true colour (RGB) low-spectral/high-spatial resolution and another high-spectral/low-spatial resolution. The proposed technique, which will be thoroughly described in what follows, is based on the automatic prior separation of urban and rural targets through the well-known Central Limit Theorem (CLT). The above mentioned prior identification relies on the fact that most rural or natural elements (i.e., fields, forests, soils, rocks) present classes with normal (Gaussian) probability density function, whereas urban classes (i.e., buildings, roads, and other city structures) generally present indescribable probability density function (Billingsley, 1995). Once the two primary targets were parameterized, a maximum likelihood classifier can be used to identify urban and rural pixels in the low spatial resolution image available. Then, in an adaptive fashion, our strategy follows by assigning pre-identified urban areas to classification using high spatial resolution image data associated with object-based analysis, whereas rural areas are assigned with low spatial resolution images and associated to pixel-level analysis.

The combination of two simultaneous approaches and images of different spatial and spectral resolutions aims at producing a robust tool for optimizing classification of the scenes covering complex and heterogeneous environments by mainly two important reasons: (1) consideration of multi-source data, attributes, and classification strategy for separate areas, and (2) limiting the number of classes for each classification problem, reducing overlap among classes.

Materials and methods

First stage: prior identification of urban and rural areas

The hybrid classification proposed here assumes the existence of two images covering the same region: a low spatial resolution multispectral image, which will be used to classify the rural part of the scene, and a high spatial resolution (not necessarily multispectral) to classify the urban portion of the scene. For rural areas, it is very convenient to have a multispectral image, with as many channels as it needs for a correct characterization of the natural targets involved (Fisher et al., 2017). This assumption relies on the fact that the chemical composition is very important to characterize the subtle differences among spectral signatures. At the same time, for the urban area, the mandatory image parameter is the pixel size, not the number of channels (Myint, Gober, Brazel, Grossman-Clarke, & Weng, 2011). For example, for buildings and roads, the shape and texture of objects are crucial information for recognizing them.

We assume images with fine spatial registration (geometrical alignment) and radiometrically
corrected. The first stage is the core of the proposed technique. It uses the low spatial resolution image to separate the hereafter called primary classes: urban $\omega_U$ and rural $\omega_R$. It is important to stress that the aim at this stage is not to find the definitive classes of targets, but only to identify the nature of targets on the scene, whether urban or rural. We understand as rural those classes representing natural elements like waterbodies, fields, rocks, bare soils, forests, crops, etc. Therefore, even in the early stages, rural class $\omega_R$ can assume more than one subclass $j$.

We employ the Central Limit Theorem (CLT) to perform the task of separating $\omega_U$ and $\omega_R$ by considering these to primary classes as two different populations. CLT states that, given a set of sufficiently large samples collected from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population (Zongh et al., 2017). Furthermore, the set of samples will approximate to a Normal/Gaussian distribution pattern, with variances approximately equal to the variance of the whole population divided by each sample's size $n$. Following this theorem, images with sufficiently large pixel sizes (low spatial resolution data) including contributions from many targets together are expected to have classes presenting Normal/Gaussian probability density distributions, which allow us to determine, in a priori fashion, the main nature of the pixels: whether urban or rural.

To better understand the proposed method and the adequacy of CLT to the related problem, let $x_1, x_2, \ldots, x_n$ be a randomly selected sample of size $n$, a sequence of independent and identically distributed random variables drawn from distributions of expected values given by $\mu$ and finite variances given by $\sigma^2$. Consider we are interested in the samples average $X_n$ of these random variables.

$$X_n = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1}$$

The theorem implies that sample averages converge to the expected value $\mu$ as $n \to \infty$ and, for large enough size $n$, the distribution of $X_n$ is closer to the Normal distribution with mean $\mu$ and variance $\sigma^2/n$. As $n$ approaches to infinity, the random variables $\sqrt{n}(X_n - \mu)$ converge to a normal distribution with $\mu = 0$, $N(0, \sigma^2)$.

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} x_i - \mu \right) \xrightarrow{d} N(0, \sigma^2) \tag{2}$$

Thus, spectral generalization caused by pixels with large sizes is important and contributes to the proper operation of the proposed framework. To adapt the CLT for the real problem approached here, we transfer the concept of one-dimensional sample average $X_n$ to the multispectral response $r_i$ of each pixel over the low spatial resolution image. We assume each pixel's response $r_i$ as a linear combination of targets included inside it (Blaschke, Lang, & Hay, 2008), where $\{p_1, p_2, \ldots, p_m\}$ are the proportions occupied by $m$ targets. Then, the spectral response $r_k$ for each channel $k$ can be depicted as:

$$r_k = r_{k1}p_1 + r_{k2}p_2 + \ldots + r_{km}p_m \tag{3}$$

where $r_k$ is the spectral response of the pixel in the $k$th spectral band, $r_{km}$ is the pure spectral response of each target present in the pixel in the $k$th spectral band.

The above demonstration shows the multi-source nature of large pixels under images including varied targets. Exploratory experiments have indeed proved our initial assumption regarding the expected Normal distributions of primary classes included in low spatial resolution images. Since the aim of the proposed method is not to estimate the proportions of pixel's compounds (unmixing), the number and nature of targets in each pixel along the image can vary, but causing any loss of validity.

The proposed method proceeds by collecting samples of pixels corresponding to the primary classes directly over the low spatial resolution image (i.e., urban $\omega_U$ and as many rural as exists $\omega_R$). The statistical information with normal distribution assumed for these classes (e.g., vector means $\mu_c$ and covariance matrix $\Sigma_c$) can then be derived and used to feed parametric classification rules. As stated before, data presenting this behaviour show a high level of differentiation by probabilistic classifiers, such as maximum likelihood, which can be expressed by the following membership function (Eq.4), derived from Bayes theorem:

$$\Phi_c(x_i) = P(x_i|\omega_c) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\det{\Sigma_c}}} \exp\left(-\frac{1}{2}(x_i - \mu_c)^T\Sigma_c^{-1}(x_i - \mu_c)\right) \tag{4}$$

where $\Phi_c(x_i)$ is the probability density function of a pixel $x_i$ belonging to class $\omega_c$, which can initially assume urban ($\omega_U$) or rural ($\omega_R$), $\Sigma$ is the dimensionality of the data, $\Sigma_c$ the spectral response of pixel $i$, $\mu_c$ is the mean vector and $\Sigma_c$ the covariance matrix of class $\omega_c$. These samples are then used to train the supervised classifier, which will later be used to determine the primary classes over the entire scene studied. The expected result is a mask separating rural and urban zone able to direct what kind of classifier is applied in each area. The suggested mask can be built from the following rule:

$$M(i) \in \omega_c \text{ if } \Phi_c(x_i) = \text{max}\{\Phi_c(x_i)\} \forall c \in \{U, R\} \tag{5}$$

where $M(i)$ corresponds to the position of the pixel $x_i$ in the mask $M$ and $\Phi_c(x_i)$ is the membership vector containing the $\Phi_c(x_i)$ values for each class. The urban ($\omega_U$) and rural ($\omega_R$) classes proceed to the second stage of the methodology. At this point, the user can
consider performing a morphologic dilation of few pixels along the urban area perimeter to guarantee to enclose of urban targets. The rationale behind this procedure is the assumption that it is even worth including rural classes inside the urban area by mistake, instead of letting urban elements outside it to be wrongly classified as rural.

Second stage: adaptive classification system

As the literature suggests, due to the high-frequency spectral behavior verified in urban areas, these sites have shown better classification results when classified by specialized techniques (Zanotta, Haertel, Shimabukuro, & Renno, 2014), which are able to take into account many parameters and specificities of targets (Lu, Hetrick, & Moran, 2010). At the same time, to prospect the ability to handle information from many kinds simultaneously and to avoid multilabelling of unique objects formed by groups of pixels, the most recent studies have suggested using object-based approaches for urban classification (Blaschke et al., 2008). Conversely, rural environments are more appropriately classified using detailed multispectral information, instead of data about the shape, texture, or spatial context of targets. Therefore, the large is the number of available spectral channels, the better is the recognition of the target. Many kinds of land cover classes like vegetation, rocks, and soils present very similar characteristics, which are often differentiated only by detailed inspection of spectral signatures (Dinis et al., 2010).

The portion of the high spatial resolution image \( M \) recognized as the urban area is then segmented and directed to the complementary classification step. The segmentation process intends the aggregation of similar neighboring pixels to produce objects with improved attributes (Jensen & Lulla, 1987). For sake of simplicity, we chose the widely used region growing segmentation technique available in many image processing packages. Region growing starts by merging individual pixels using spectral similarity criteria, which can be more adequately classified by using not only their spectral attributes, but also texture, shape, and spatial context features (Blaschke, Kelly, & Merschdorf, 2015). Resulting objects are then classified using one of the techniques suitable for urban environments. The most popular approaches for this type of application are those which can handle many classes at the same time, while avoiding overfitting and making optimized usage of the available attributes. Some modern examples are the hierarchical Random Forests (RF) (Jiang, Wang, Yang, Xie, & Cheng, 2010), since it can manage different attributes according to the importance of each one to the specific problem addressed. Traditional decision trees classifiers tend to learn highly irregular patterns, which frequently causes overfitting of the training samples. RF is an alternative that avoids overfitting by averages multiple deep decision trees trained in different parts of the same training dataset (Hastie, Tibshirani, & Friedman, 2009).

The area in \( M \) recognized as rural can keep the original classes received at the primary stage, or can be classified again using the low spatial resolution image by a pixel-based approach and generalist classification technique, such as Linear or Quadratic Discriminant Analysis (LDA, QDA) or Maximum Likelihood. The generalist classification technique to operate on the low spatial resolution image is defined as \( G \), whereas the classification technique chosen to operate on the segmented high spatial resolution image is defined as \( H \). The adaptive classification of the entire scene proceeds obeying the following rule:

\[
\begin{align*}
(\mathbf{x}_i & \Rightarrow H \text{ if } M(i) \in \omega_U) \\
(\mathbf{x}_i & \Rightarrow G \text{ if } M(i) \in \omega_R)
\end{align*}
\]  

(6)

where one pixel \( \mathbf{x}_i \) is expected to be classified by high spatial resolution image, as well as technique \( H \), only if it is recognized as urban area in the first stage \( M(i) \in \omega_U \). Conversely, if the pixel \( \mathbf{x}_i \) is recognized as rural in the first step \( M(i) \in \omega_R \), then this element is expected to be classified by the low spatial resolution data, operated by technique \( G \). A flowchart of the proposed technique is presented in Figure 1.

The resulting classification map is expected to present an improvement according to traditional methods applying a single rule throughout the scene, ignoring the fact that it contains targets of distinct natures. As said before, the core of the proposed technique is to automatically exploit the advantages of each source of data and the potentials of proper classification tools for every specific environment. Furthermore, the reduction in the number of classes available for each classification system is expected to avoid overlaps and confusion among classes, improving overall classification results.

Results

Data description

In order to test and exemplify the performance of the methodology suggested in this study, we performed one experiment with an area located in Cape Town, Western Cape Province, South Africa. The images are geometrically and radiometrically/atmospherically corrected. Standing for low spatial resolution data we have a Landsat 8-OLI, acquired on 3 September 2013 (Figure 2(a)). The image has 30 m spatial resolution for the spectral bands used in the experiment (1–7). The high spatial resolution image came from GeoEye-1 acquired on 31 July 2013, with 1.65 m spatial resolution (Figure 2(b)). Two images with different resolutions
covering the same area were used: one low spatial resolution multispectral image covering the whole study area, and one high spatial resolution image covering at least the urban spots.

**Validation dataset**

Reliable ground truth was prepared by expert visual interpretation to allow accuracy assessment, which was made in two steps: first, based only on the low spatial resolution image to test prior identification of primary targets by CLT (rural targets and generic urban), and second, based on the low and high spatial resolution images simultaneously to assess the final classification result (rural and urban targets in fine detail). It is important to stress that the second ground truth was built by drawing only suburban classes (roofs, roads, trees, soils, etc.) directly over the high-resolution image, but avoiding areas considered as rural according to the first ground truth data. Then, this second ground truth related only to the detailed urban area was merged to the first (only rural areas) in order to produce the absolute ground truth, which was finally used to assess the performance of the entire classification. Ambiguous areas (black areas) were not labelled and consequently disregarded during the accuracy assessment.

**Experiment with Landsat 8-OLI combined to Geoeye-1**

The selected area includes rocks surrounded by vegetation and some portions of the urban area spread. Primary samples of forest, field, rocks, and urban areas were collected directly on the image. Based on the CLT (Zhong et al., 2017), the image received the primary
classification by maximum likelihood considering Normal statistical distribution of classes to separate the primary targets selected on the scene and using OLI-Landsat 8. The initial supposition of Normal distribution of primary classes was confirmed by analyzing the histograms of Figure 3, which also over plots the estimated probability density functions (dotted lines). For this experiment, bands 4 and 7 were sufficient to separate all the four primary classes.

The resulting mask \( M(i) \) separating the primary classes is showed in Figure 4(a). As can be seen, considering the trade-off mentioned at the end of section 2, we have performed a morphologic dilation of few pixels along the perimeter of the urban area to expand it (rounded features at the edges of the urban area).

Proceeding to the second stage of the method (adaptive classification), the high-resolution GeoEye-1 image was segmented by a basic region growing technique only over the identified urban regions (grey colour in Figure 4(a)), and then classified using RF, the technique selected in this experiment to operate on the high-resolution image. The areas identified as rural (forest, grass, and rocks) were kept with the class resulting from the first stage.

Therefore, representative samples of roofs (clay, concrete, and fibrocement), forests and paved roads were collected over the high-resolution GeoEye-1 image and used for training purposes. The RF classifier was trained by using the C4.5 algorithm (Jensen & Lulla, 1987) with the available spectral image features. Finally, the classification maps were merged to produce one final image for each classifier RF (Figure 4(c)).

### Analysis

The classification map resulting from the first stage (Figure 4(a)) was validated using the reference data built by vectorization from the Landsat 8-OLI image, shown in Figure 4(b). It is important to stress that this initial map aimed to test only the ability to separate between rural (forest, grass, and rocks) and urban area in terms of overall and average accuracies. This result retrieved an overall accuracy of 82.9% and an average accuracy of 84.0%. Qualitatively, it is also possible to notice spatial correspondence between Figure 4(a,b).

Most importantly, the pre identification of the urban area resulted in a high classification confidence, which was greatly aided by the post-classification morphological dilation process. It is worth noting that, to avoid further classification errors from the first to the second stage, it is preferable to obtain an excess than an absence of an urban area resulting from the first stage.

The final classification map was then obtained by refining the classification of the urban area through the high-resolution image (GeoEye-1). The end result was validated using the fully ground truth present in Figure 2c. The performance of the proposed methodology was compared with the traditional classification, i.e., using only a single image and one classification technique. To allow comparison, we selected the same RF classification technique used to test the proposed technique, as well as an identical set of class samples. The results obtained in terms of overall and average accuracies are presented in Table 1. We see that for overall accuracy the scores achieved for the proposed method were significantly higher than those found by the traditional approach. Tables 2 and 3 present the confusion matrices computed for both tested techniques. Figure 5 shows details in the urban area of the classified images compared to the same parts in the Geoeye image.

We can also visually verify that the changes provided by the proposed approach caused punctual increases in the classification performance over the entire image and for all classes. However, due to the large difference between the number of pixels for each

---

**Figure 3.** Primary classes histograms and corresponding probability density functions (dotted lines) for channels 4 and 7. The result shows urban class can be adequately separated in the first stage from the remaining rural classes by means of both combined channels.
class, the most important measure in this scenario is the average accuracy, which was also higher for the proposed method. We can also see through the maps of Figure 4 that much of this result was due to the generalization caused by the classification at the first stage, when providing fine separation between forest

![Figure 4](image-url)

**Figure 4.** (a): Primary classification mask resulting from the first stage, (b): Reference map for primary classes based on the low spatial resolution image by visual interpretation (ground truth first stage), (c): Final map using proposed adaptive classification and RF method for urban area, (d): Traditional object-based classification using only the Geoeye-1 image and RF method.

| Table 1. Average and overall classification accuracies (%) measured using RF for the traditional classification paradigm and the proposed simultaneous methodology. |
| --- |
| **Method** | RF Traditional | RF Simultaneous |
| Overall | 56.8 | 81.0 |
| Average | 54.1 | 67.7 |

| Table 2. Confusion matrix regarding experiment applying the proposed hybrid method. |
| --- |
| **Classified** | Grass | Forest | Soil | Asphalt | Clay roof | Concrete roof | Fibrocement roof | **Σ (row total)** |
| Grass | 46,141 | 3092 | 2884 | 787 | 2 | 0 | 0 | 57,921 |
| Forest | 14,504 | 58,435 | 204 | 3262 | 103 | 403 | 87 | 76,998 |
| Soil | 2568 | 129 | 27,516 | 578 | 0 | 0 | 0 | 30,791 |
| Asphalt | 1005 | 2789 | 0 | 557 | 44 | 367 | 138 | 4900 |
| Clay roof | 0 | 4 | 0 | 0 | 462 | 0 | 0 | 466 |
| Concrete roof | 3 | 134 | 0 | 91 | 280 | 2551 | 126 | 3185 |
| Fibrocement roof | 0 | 6 | 0 | 0 | 0 | 278 | 103 | 387 |
| **Σ (column total)** | 64,221 | 64,589 | 30,604 | 5275 | 891 | 3614 | 361 | 169,648 |
Discussion

We can certainly expect that using only the low spatial resolution image to classify the entire area, the results over the urban area would be very poor and inaccurate due to the presence of small parcels and many kinds of materials in the cities. Conversely, by using only the high spatial resolution image over the entire region would certainly prevent the rural area to be optimally classified. This is due because rural/natural areas do need as many spectral bands as it is possible to obtain the best representation of the spectral signature of the targets, which is the key to achieve the best classification result.

The comparison of results produced by the proposed method with results generated using only the object-based classification by RF has shown that the proposed approach presented encouraging qualitative and quantitative results, especially in regions where pixel-based classifiers tend to fail: urban areas or areas with high radiometric variability and areas where the object-based approach is more suitable. Conversely, rural areas including fields, soils, minerals, rocks, and trees could be correctly classified due to the sufficient number of spectral bands available in the low spatial resolution image.

Using the predictions of CLT, the elements corresponding to primary targets could be effectively separated using the probabilistic Gaussian maximum likelihood classifier. The union of partially produced maps for each environment achieved an optimized result, matching the advantages of both methods in a single scene containing heterogeneous targets. As can be seen in Figure 5, the urban area present a similar classification provided by the traditional approach. However, the absence of rural classes in the problem could allow classification with some improvements, not showing rural targets in these areas. On the other hand, the rural area, previously

Table 3. Confusion matrix regarding experiment applying the traditional classification method (object-based).

| Classified | Grass | Forest | Soil | Asphalt | Clay roof | Concrete roof | Fibrocement roof | ∑ (row total) |
|------------|-------|--------|------|---------|-----------|---------------|------------------|--------------|
| Grass      | 4037  | 1058   | 381  | 441     | 8         | 22            | 0                | 5947         |
| Forest     | 38,938| 59,350 | 2346 | 1548    | 60        | 282           | 33               | 102,557      |
| Soil       | 20,180| 2295   | 27,829| 480     | 134       | 231           | 74               | 51,223       |
| Asphalt    | 786   | 362    | 11   | 2096    | 27        | 206           | 79               | 3567         |
| Clay roof  | 0     | 76     | 0    | 21      | 480       | 2             | 0                | 579          |
| Concrete roof | 236 | 1363   | 37   | 682     | 180       | 2439          | 85               | 5022         |
| Fibrocement roof | 44 | 85     | 0    | 7       | 2         | 432           | 183              | 753          |
| ∑ (column total) | 64,221 | 64,589 | 30,604| 5275    | 891       | 3614          | 454              | 169,648      |

Figure 5. (a): GeoEye-1 in 3 2 1 true colour composition, (b) Final map using proposed adaptive classification and RF for urban area, (c): Traditional object-based classification using only the Geoeye-1 image and RF. All figures have details depicting parts of interest located on same areas.
classified by using the multispectral image and parametric (maximum likelihood classifier) was more stable, showing less noisy like pixels, when compared to the traditional object-based approach.

Maybe one of the major drawbacks of the proposed method is the first stage, when prior separation between rural and urban targets is proceeded. This crucial point can affect greatly the final results, once some urban areas can be wrongly confused with rural. This issue has encouraged us to investigate alternative procedures to find the urban region with even more precision. Nighttime images as auxiliary data is one of the possibilities. Nighttime data register artificial light coming from the surface of the Earth, potentially indicating areas covered by urban spots. Other possibilities are fixed maps and Synthetic Aperture Radar (SAR) images.

The proposed method is a practical way for mapping heterogeneous areas reaching soundness classification results, overcoming sensor limitations regarding spatial and spectral resolution. The main advantages achieved by the proposed method include:

1. The optimization of the classification process by automatically selecting the most appropriate base image and classification technique to different areas in the same problem (high spatial resolution for urban and high-spectral resolution for rural).

2. The independence of available classes for different primary targets when the classification process is made in separate. As the rural areas do not present the diversity of classes found in urban sites, the classification tends to produce more consistent results, since each problem is posed with a separate set of samples.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Leticia Sartorio http://orcid.org/0000-0001-6936-9939
Daniel Zanotta http://orcid.org/0000-0003-2959-6525

References

Aguirre-Gutiérrez, J., Seijmonsbergen, A. C., & Duivenvoorden, J. F. (2012). Optimizing land cover classification accuracy for change detection, a combined pixel-based and object-based approach in a mountainous area in Mexico. Applied Geography, 34, 29–37.

Bhaskaran, S., Paramananda, S., & Ramnarayan, M. (2010). Per-pixel and object-oriented classification methods for mapping urban features using Ikonos satellite Data. Applied Geography, 30(4), 650–665.

Billingsley, P. (1995). Probability and measure. Wiley Series in Probability and Mathematical Statistics, Hoboken, NJ.

Blaschke, T., Kelly, M., & Merschdorf, H. (2015). Object-based image analysis: Evolution, history, state of the art, and future vision. In P. Thenkabail (Ed.), In remotely sensed data characterization, classification, and accuracies (pp. 277–294). Boca Raton, Florida: CRC Press.

Blaschke, T., Lang, S., & Hay, G. (2008). Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications. Springer Science & Business Media, Heidelberg, Berlin.

Chuvieco, E. (2016). Fundamentals of satellite remote sensing: An environmental approach. Boca Raton, Florida: CRC Press.

Dinis, J., Navarro, A., Soares, F., Santos, T., Freire, S., Fonseca, A., … Tenedô, J. (2010). Hierarchical object-based classification of dense urban areas by integrating high spatial resolution satellite images and Lidar elevation data. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 38(6), 4.

Duro, D. C., Franklin, S. E., & Dublé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using spot-5 Hrg imagery. Remote Sensing of Environment, 118, 259–272.

Ferreira, M. P., Zortea, M., Zanotta, D. C., Shimabukuro, Y. E., & Souza Filho, C. R. (2016). Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. Remote Sensing of Environment, 179, 66–78.

Fisher, J. R. B., Eileen, A., James Dennyd-Frank, P., Kroeger, T., & Boucher, T. M. (2017). Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality. Remote Sensing in Ecology and Conservation. doi:10.1002/rsa.26117

Gholoobi, M., & Kumar, L. (2015). Using object-based hierarchical classification to extract land use land cover classes from high-resolution satellite imagery in a complex urban area. Journal of Applied Remote Sensing, 9(1), 096052–096052.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. New York, NY: Springer.

Herold, M., Roberts, D. A., Gardner, M. E., & Dennison, P. E. (2004). Spectrometry for urban area remote sensing—Development and analysis of a spectral library from 350 to 2400 Nm. Remote Sensing of Environment, 91(3–4), 304–319.

Jensen, J. (2009). Remote sensing of the environment: An earth resource perspective (2th ed.). New Delhi, INDIA: Pearson Education India.

Jensen, J. R., & Lulla, K. (1987). Introductory digital image processing: A remote sensing perspective. Milton Park, UK, Taylor & Francis.

Jiang, L., Wang, W., Yang, X., Xie, N., & Cheng, Y. (2010). Classification methods of remote sensing image based on decision tree technologies, Paper presented at the International Conference on Computer and Computing Technologies in Agriculture (353–358), Nanchang, China.

Lillesand, T., Kiefer, R. W., & Chipman, J. (2014). Remote sensing and image interpretation. Hoboken, NJ: John Wiley & Sons.

Lu, D., Hetrick, S., & Moran, E. (2010). Land cover classification in a complex urban-rural landscape with Quickbird imagery. Photogrammetric Engineering & Remote Sensing, 76(10), 1159–1168. doi:10.1111/j.1112-7610.2010.01159.x

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., & Liu, Y. (2017). A review of supervised object-based land-cover image
classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 277–293.
Mather, P., & Tso, B. (2016). *Classification methods for remotely sensed data*. Boca Raton, Florida: CRC press.
Moosavi, V., Talebi, A., & Shirmohammadi, B. (2014). Producing a landslide inventory map using pixel-based and object-oriented approaches optimized by Taguchi method. *Geomorphology*, 204, 646–656.
Myint, S., Gober, P., Brazel, A., Grossman-Clarke, S., & Weng, Q. (2011). Per-pixel Vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, 115(5), 1145–1161.
Weih, R. C., & Riggan, N. D. (2010). Object-based classification Vs. Pixel-based classification: Comparative importance of multi-resolution imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(4).
Whiteside, T. G., Boggs, G. S., & Maier, S. W. (2011). Comparing object-based and pixel-based classifications for mapping savannas. *International Journal of Applied Earth Observation and Geoinformation*, 13(6), 884–893.

Zanotta, D. C., Haertel, V., Shimabukuro, Y. E., & Renno, C. D. (2014). Linear spectral mixing model for identifying potential missing endmembers in spectral mixture analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5), 3005–3012.
Zhao, W., Du, S., & Emery, W. J. (2017). Object-based convolutional neural network for high-resolution imagery classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(7), 3386–3396.
Zhong, Y., Fei, F., Liu, Y., Zhao, B., Jiao, H., & Zhang, L. (2017). Satcnn: Satellite image dataset classification using Agile convolutional neural networks. *Remote Sensing Letters*, 8(2), 136–145.
Zhong, Y., Ma, A., Ong, Y., Zhu, Z., & Zhang, L. (2017). Computational intelligence in optical remote sensing image processing. *Applied Soft Computing*, 64, 75–93.
Zhou, W., Troy, A., & Grove, M. (2008). Object-based land cover classification and change analysis in the Baltimore metropolitan area using multitemporal high resolution remote sensing data. *Sensors*, 8(3), 1613–1636.