



HAL
open science

Interactive Multiscale Classification of High-Resolution Remote Sensing Images

Jefersson Alex dos Santos, Philippe-Henri Gosselin, Sylvie Philipp-Foliguet,
Ricardo da Silva Torres, Alexandre Xavier Falcao

► **To cite this version:**

Jefersson Alex dos Santos, Philippe-Henri Gosselin, Sylvie Philipp-Foliguet, Ricardo da Silva Torres, Alexandre Xavier Falcao. Interactive Multiscale Classification of High-Resolution Remote Sensing Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2013, PP (99), pp.1-15. hal-00799071

HAL Id: hal-00799071

<https://hal.science/hal-00799071>

Submitted on 11 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interactive Multiscale Classification of High-Resolution Remote Sensing Images

Jefersson Alex dos Santos, Philippe-Henri Gosselin, Sylvie Philipp-Foliguet,
Ricardo da S. Torres and Alexandre Xavier Falcão

Abstract

The use of remote sensing images (RSIs) as a source of information in agribusiness applications is very common. In those applications, it is fundamental to identify and understand trends and patterns in space occupation. However, the identification and recognition of crop regions in remote sensing images are not trivial tasks yet. In high resolution image analysis and recognition, many of the problems are related to the representation scale of the data, and to both the size and the representativeness of the training set. In this paper, we propose a method for interactive classification of remote sensing images considering multiscale segmentation. Our aim is to improve the selection of training samples using the features from the most appropriate scales of representation. We use a boosting-based active learning strategy to select regions at various scales for user's relevance feedback. The idea is to select the regions that are closer to the border that separates both target classes: relevant and non-relevant regions. The experimental results showed that the combination of scales produces better results than isolated scales in a relevance feedback process. Furthermore, the interactive method achieved good results with few user interactions. The proposed method needs only a small portion of the training set to build classifiers that are as strong as the ones generated by a supervised method that uses the whole training set.

Index Terms

Interactive Classification, Multiscale Classification, Boosting, Support Vector Machines, Active Learning.

I. INTRODUCTION

The use of remote sensing images (RSIs) as a source of information in agribusiness applications is very common. In those applications, it is essential to identify and understand trends and patterns in space

J. A. dos Santos, R. da S. Torres and A. X. Falcão are with the Institute of Computing, University of Campinas, Brazil email: jsantos@ic.unicamp.br

J. A. dos Santos, P. Gosselin and S.Philipp-Foliguet are with ETIS, CNRS, ENSEA, University of Cergy-Pontoise, France

occupation. However, the identification and recognition of crop regions in remote sensing images are not trivial tasks yet.

Apart from the typical RSI classification problems, there are some specific issues related to the agriculture domain. For instance, as far as the identification of coffee crop regions is concerned, there are additional challenges to be faced. Coffee crops are usually cultivated in mountainous regions (for example, in Brazil) [1]. This fact often causes shadows and distortions in the spectral information. Difficulty arises in the classification and interpretation of shaded objects due to the reduction of their spectral information [2]. Moreover, the growing of coffee is not a seasonal activity, and, therefore, in the same region, there may be coffee plantations of different ages. That also affects the observed spectral patterns.

Typically, the RSI classification process uses supervised learning, which comprises three main steps: data representation, feature extraction, and training. Data representation indicates the objects for classification (e.g., pixels [3], blocks of pixels [1], regions [4], and hierarchy of regions [5]). Feature extraction provides a mathematical description for each object (by taking into account, for example, spectral characteristics, texture, shape). Training learns how to separate objects from distinct classes by building a classifier based on machine learning techniques (for instance, support vector machines [6], optimum path forest [7], genetic programming [1], and neural networks [8]).

Concerning data representation, pixel-based methods have always been very popular for RSI classification [9]. They only use the value of the pixel in each band as a spectral signature to perform the classification. Indeed, concerning hyperspectral images, it is possible to associate a detailed spectral signature with each pixel, whose dimensions usually correspond to large areas. Therefore, the pixel-based analysis seems to be enough to satisfy the requirements of representation and feature extraction necessary to perform classification [3]. However, to the extent that high-resolution images became available to the civil community, new approaches of representation and feature extraction have been proposed in order to make better use of data [10]. Several methods using region-based analysis presented improvements in results when compared with traditional methods based on pixels [2, 11–15]. The main problem of region-based approaches is their dependence on good algorithms of segmentation. Moreover, defining the most appropriate scale of segmentation is still a challenging research topic in remote sensing area [5, 16–18]. Some approaches have been proposed to overcome this problem by selecting regions of different sizes during segmentation [6]. Other methods exploit data from different scales [17–22], or select a single scale from a hierarchy of segmented regions [5]. A particular case of region-based strategy is the plot-

based approach [23–25], which exploits cartographic information to define region boundaries. The main advantage is that the use of cartographic data enables a better delineation of user's interest objects than automatic segmentation techniques. Its main problem is the lack of available cartographic data. Thus, most of the papers related to the plot-based approach is focused on urban applications.

Regarding the training process, there are many research challenges that concern the labeling of samples. The main important ones are related to the size and redundancy of the training set [26]. The size and quality of the training set have a direct impact on the execution time needed for training and on the final result of the classification. In addition, labeling of samples often requires visits to the study site, which can add extra costs to the analysis. The training set must, thus, be carefully chosen, avoiding redundancy patterns, but also ensuring a good representation of the considered classes. In order to assist users in selecting samples, several interactive methods have been proposed for dealing with remote sensing data [7, 27–33].

To a limited extent, there are methods, in the literature, that propose interactive approaches or active learning strategies to address the challenges of defining training samples for remote sensing classification. Nevertheless, it is worth mentioning that none of the above cited methods have been proposed to work with features extracted from segmented regions. Moreover, as far as we know, there are no proposed region-based classification methods that address the problem of selecting samples for training.

In this paper, we propose a method for interactive classification of remote sensing images considering multiscale segmentation. Our aim is to improve the selection of training samples using the features from the most appropriate scales of representation. Figure 1 gives an overview of the architecture used in our approach for interactive classification. This kind of architecture is very common in information retrieval systems with relevance feedback [7, 34–37]. The framework is composed by three main processing modules: segmentation, feature extraction, and classification. Segmentation and feature extraction are offline steps. When an image is inserted into the system, the segmentation is performed, building a hierarchical representation of regions. Feature vectors from these regions are then computed and stored.

The interactive classification starts with the user's annotation. He/she selects a small set of relevant and non-relevant pixels. Using these pixels as training set, the method builds a classifier to label the remaining pixels. Although the training set is at the pixel level, the training is performed by using features extracted from the segmented regions for each considered scale. At the end of the classification step, the method selects regions for possible feedback. When the result of the classification is displayed, the user feeds the

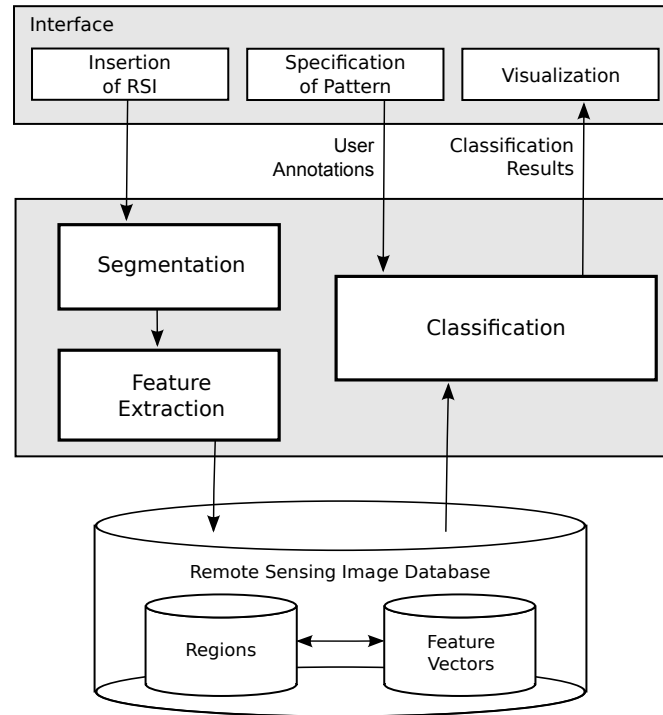


Fig. 1. Architecture of the interactive classification system.

system by labeling the region with the correct class. These steps are repeated until the user finishes the process. The final classification is a multiscale result combining all scales of segmentation.

Our method uses a hierarchical segmentation of regions, from the pixel level to the entire image. For the training stage, we propose a kind of boost-classifier adapted to the segmentation, which takes advantage of various region features. In each iteration, this method builds a strong classifier from a set of weak ones. The weak classifiers are SVMs (Support Vector Machines) with a linear kernel, each trained for one feature descriptor of one scale of segmentation. We use a boosting-based active learning strategy to select regions for user's relevance feedback. The idea is to select the closest regions to the border separating both target classes. Those regions are theoretically the most uncertain regions.

Our method differs from the other classification approaches in several aspects. The main contributions are:

- 1) **An innovative interactive method for classifying segmented regions.** It allows the user to refine the classification results by changing the scale of representation from the coarsest to the finest ones along the iterations. This solves the best-scale definition problem, since all the scales are involved.
- 2) **A strategy to reduce the multiscale training time.** The proposed hierarchical multiscale training considers a small set for each training scale, which is desirable in an interactive learning process.

- 3) **A strategy for selecting segmented regions.** We proposed a strategy based on active learning to select regions for user annotation by exploiting the output from our boosting-based algorithm.
- 4) **A feature combination strategy.** Using several weak classifiers, our method allows the combination of different types of descriptors (such as color, texture, and shape). Also, our method allows the improvement of the combination insofar as the most significant samples are added to the training set.

Experimental results show that the combination of scales produces better results than isolated scales in a relevance feedback process. The interactive method achieves good results with few iterations. Furthermore, by using only 5% of the pixels of the training set, our approach can build classifiers which are as strong as the ones generated by a supervised method using the whole training set [18].

This paper is outlined as follows. Section II presents related work. Section III presents some concepts related to representation and description necessary to understand the proposed approach. Section IV introduces our method for multiscale training and classification. Experimental results are presented in Section V. In Section VI, we present our conclusion and discuss future work.

II. RELATED WORK

A study of published works between 1989 and 2003 examined the results and implications of RSI classification research [9]. According to this study, despite the high number of approaches in that period, there was not significant improvement in terms of classification results. Most of the proposed methods were pixel-based. These methods try to estimate the probability of each pixel to belong to the possible classes employing statistic measures based on spectral properties. The *Maximum Likelihood Classification* (MLC) [3] has remained as one of the most popular methods for RSI classification.

The improvements in sensor technologies increased the accessibility to high-resolution and hyperspectral imagery. As a result, new approaches were developed to make better use of the available data [10]. Two main research approaches to address those issues can be observed in the literature. The first one, which is more related to high-resolution images, focuses on data representation and feature extraction [6, 8, 11–13, 21, 22]. The other approach, more associated with pixel-based classification methods, is focused on issues related to the selection of samples for training and the inclusion of the user in the classification process [26–29, 31–33]. In the next two subsections, we discuss each of these approaches. Concerning the first one, we highlight proposed methods for classification based on regions. Regarding the second approach, we point out proposed techniques related to interactive classification of remote sensing images.

A. Region-based Classification Methods

Initially, advances towards the classification of high-resolution data focused on the use of the neighborhood of the pixels in the analysis, which included texture descriptors [10].

More recently, many studies [11–13] have considered information encoded in regions (group of pixels) for RSI classification tasks. The growth of classification approaches based on regions has been analyzed in [38]. According to Blaschke et al., the goal of OBIA (*Object Based Image Analysis*) is to outline objects that are useful in the images. It combines, at the same time, image processing and features of Geographic Information Systems (GIS) aiming to use spectral and contextual information seamlessly. The paper shows that the growth in the number of new approaches is accompanied by the increase of the accessibility to high-resolution images and, hence, the development of alternative techniques to the classification based on pixels. As pointed out by the authors, the growth in research involving OBIA was motivated in part by the use of commercial software eCognition [4]. The software has allowed research involving classification of regions, enabling the inclusion of data from different scales.

These new trends have encouraged research studies that compare techniques based on pixels and/or regions [2, 12, 14, 15], and propose new segmentation techniques that support the classification of regions in RSIs [20, 39–41]. Likewise, new research has been carried out to take advantage of the use of multiple scales of data [6, 8, 21, 22].

B. Interactive Classification Methods

Several recent approaches handle the RSI classification problem by exploiting the user interactions [7, 27–30]. The main purpose of these methods is to help the user to build a representative training set, improving classification results during the iteration. According to Tuia et al. [26], in high-resolution imagery, the selection of training data can be easily done on the image. However, several neighboring pixels can be included in the selection, carrying the same spectral information. Consequently, the training set may be highly redundant. Furthermore, the labeling of training samples may require visits to the studied places, as those samples may be linked to geographical references. That adds extra costs to the classification process.

Most of the proposed methods are SVM-based [28, 29, 31, 32]. In these approaches, active learning plays a key role. It provides an interactive way to build training sets that correctly represent the boundaries of separation between classes, avoiding redundancies. Rajan et al. [27], on the other hand, proposed an

approach based on active learning that can be applied to any classifier since this classifier is able to work with decision bounds. They apply the principle of selecting data points that most change the existing belief in class distributions.

Santos et al. [7, 30] have also recently proposed two interactive methods for classification of RSIs. In [30], they proposed an interactive framework based on relevance feedback, called GP_{SR} . That framework allows the classification of RSIs and the combination of distances from feature descriptors by using genetic programming (GP). In [7], they propose a new framework ($GOPF$) that integrates the Optimum-Path Forest classifier [42] and GP to perform interactive classification combining different types of features.

III. IMAGE REPRESENTATION AND FEATURE EXTRACTION

Lately, many multi-scale segmentation methods have been proposed for remote sensing purposes [16, 20, 39–41, 43, 44]. In this work, we use the scale-set representation introduced by Guigues et al. [45]. This representation is a hierarchy of regions of partitions obtained at all scales, from the pixel level until the complete image. Basically, we use the Guigues' approach because it is hierarchical (essential for our proposal) and it has a strong theoretical foundation. Anyway, the proposed approach for interactive multiscale classification is general and can exploit any other hierarchical region-based segmentation method.

Let image I be defined over a domain \mathcal{D} , a partition P is a division of \mathcal{D} into separate regions. A partition P_2 is finer than a partition P_1 if each region R of P_2 is included in one and only one region of P_1 . The scale-set representation consists in defining a set of partitions P_λ of \mathcal{D} , indexed by a scale parameter λ , such that if $\lambda_1 \leq \lambda_2$ then P_2 is finer than P_1 . The transition between P_i and P_{i+1} is obtained by merging some adjacent regions of P_i into larger regions by optimizing a criterion. The criterion we use corresponds to Mumford-Shah energy [46], which approximates the color image by a piecewise constant function, while minimizing the edge lengths. The compromise between both constraints is defined by the parameter λ . For small values of λ , the image is over-segmented, the approximation of each region by a constant is perfect, but the total length of all edges is very large. On the contrary, when λ is large, the partition contains few regions (until only one), then the approximation of each region by a constant is poor, but the total length of all edges is very small. The set of partitions has a structure of a hierarchy H of regions: two elements of H , which are not disjoint, are nested. A partition P_λ is composed by the set of regions obtained from a cut in the hierarchy H at scale λ (see Figure 2). Guigues et al. showed that

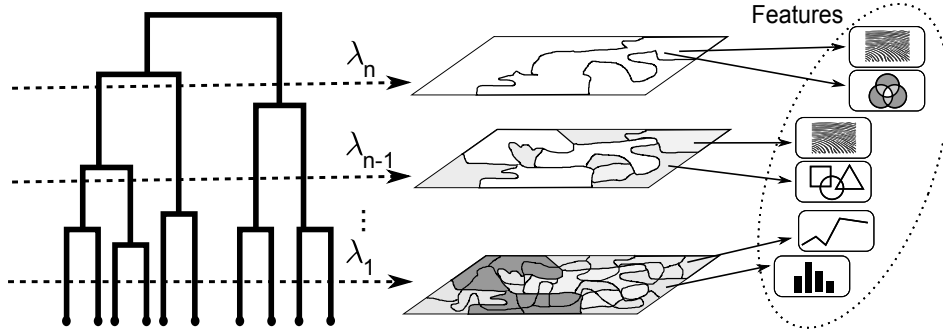


Fig. 2. Feature extraction. For each selected scale, a set of features is extracted from all regions.

this algorithm can be performed with the worst case complexity in $O(N^2 \log N)$, where N is the size of the initial over-segmentation.

Guigues algorithm is a merging process, which iteratively merges neighbouring regions by minimizing an energy criterion. It starts at pixel level, or after a watershed process, aiming to obtain regions more reliable to compute the energy. It stops when all regions are merged.

To automatically select partitions at different scales, Guigues et al. propose the use of a dichotomous cutoff-based strategy, which consists of successively splitting the hierarchy of regions in two. Each division is a dichotomous cut and creates a partition at the defined scale. Let Λ be the maximum scale in the hierarchy H , i.e., the one in which the image I is represented by a single region, the scale of cut λ is defined by $\lambda = \Lambda/2^n$, where n is the order of each division in the hierarchy.

In this work, we use the descriptor notion proposed in [47]. Figure 2 illustrates the process of feature extraction from the hierarchy. We select some scale partitions to extract features based on the dichotomous cutoff-based strategy. For each region in the partitions we use a set of descriptors to extract features. More details about the feature extraction are given in Section V-A2.

IV. THE PROPOSED INTERACTIVE CLASSIFICATION METHOD

At the end of the off-line process, the image is represented by a set of several nested partitions (in our experiments, there were 5 of them). Each region of each scale is described by a feature vector. The classification step is performed on-line, through an interactive process of region classification. The aim of RSI classification is to build a classification function $F(p)$ that returns a classification score (+1 for samples of interest, and -1 otherwise) for each pixel p of an RSI. Let us note that, even if the classification returns a result at a pixel level, the decision is based on regions of different scales that contain the pixel.

Given a set of labeled pixels as training set Y_0 and features extracted from regions of various scales, our method aims to produce a classifier to label the remaining pixels. Moreover, the method uses an active learning strategy based on user interaction to increase the training set and, hence, the classification results. Algorithm 1, which presents the proposed interactive classification process, will be further explained in the next section.

Algorithm 1 The interactive classification process.

```

1 Annotation of the initial training set  $Y_0 = \text{label of pixels}$  (see Section IV-D)
2 Build a classifier  $F_0(p)$  using multi-scale training (see Section IV-A)
3 Classify image  $I$  by using  $F_0(p)$ 
4  $i \leftarrow 1$ 
5 while user is not satisfied do
6     Select uncertain regions  $Q_i$  in the classified image (see Section IV-C)
7     Annotation of the selected regions  $Q_i$  (see Section IV-D)
8     Update the training set  $Y_i \leftarrow Y_{i-1} \cup Q_i$ 
9     Build a classifier  $F_i(p)$  using multi-scale training (see Section IV-A)
10    Classify image  $I$  by using  $F_i(p)$ 
11     $i \leftarrow i + 1$ 
12 end while

```

Algorithm 1 starts the process with the definition of the training set Y_0 annotated by the user (line 1). We consider that, in a real scenario, the samples indicated by the user may not always be representative. The training set is used to build a multi-scale classifier $F_0(p)$ (line 2). This approach is based on the boosting of weak classifiers [48]. The multi-scale classifier $F_0(p)$ is used to classify the whole image I (line 3). The feedback process starts using this initial classification result. In the loop, the user can stop the classification process or continue the classification refinement process (line 5). To select regions displayed for user annotation, also known as active learning, we exploit the notion of separating border in AdaBoost, which is originally proposed in [49]. In the refinement iterations, the following steps are performed: selection of the most uncertain regions in each scale λ (line 6); annotation of the selected regions by the user (line 7); update of the training set by adding the new labeled regions to Y_i (line 8); multi-scale training by using the new set Y_i (line 9); reclassification of the whole image I by using $F_i(p)$ (line 10).

The proposed approach is designed to assist specialists, who are our final users. Our approach expects the user to have reasonable knowledge about the region and the targets of interest. A way to create a stopping criterion is to define some validation points (it can be pixels) in the image as usually done by experts to assess the quality of supervised classification in practical situations. A validation point is a

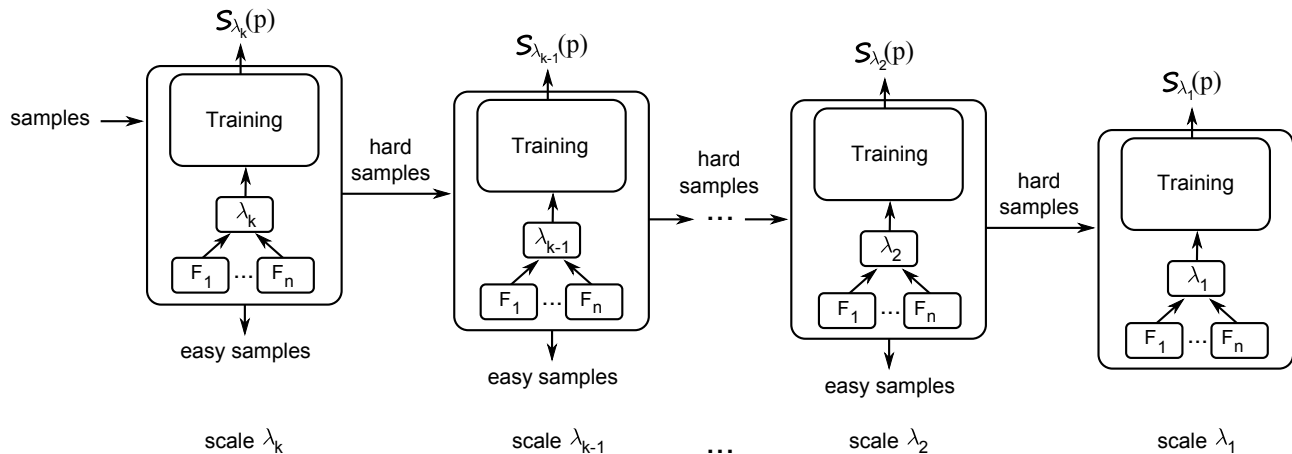


Fig. 3. The multi-scale training strategy. Adapted from [18].

well-known place in the scene which is not used to train and can be used to evaluate classification results. When the method achieves acceptable accuracy in the validation points, the user can stop the interactive process. Another option is to previously determine a number of iterations. It is important to clarify that, besides seeing the regions selected for annotation, the user can also check the classification results.

We explain in details each step of the process in the following sections. In Section IV-A, we present the multi-scale classification based on boosting. In Section IV-B, we describe the linear SVM-based configuration used as weak classifier. The active learning process is explained in Section IV-C. In Section IV-D, we present how user annotation is carried out.

A. Multi-Scale Training/Classification

The *hierarchical multi-scale classifier* (HMSC) aims at assigning a label (+1 for relevant class, and -1 otherwise) to each pixel of P_0 , taking advantage of various features computed on regions of various levels of the hierarchy. To build multi-scale classifiers, we propose a learning strategy based on boosting of weak learners.

The proposed strategy is illustrated in Figure 3. It consists of individually selecting the weak classifiers for each scale, starting from the coarsest to the finest ones. Thereby, each scale provides a different stage of training. At the end of each stage, only the most difficult samples are selected, limiting the training samples used in the next stage/scale.

In each stage/scale, the proposed method repeatedly calls *weak learners* in a series of rounds $t = 1, \dots, T$. Each weak learner creates a weak classifier that decreases the expected classification error of the

combination. The algorithm then selects the weak classifier that most decreases the error.

For each scale λ , the weak learner produces a set \mathcal{S}_λ of weak classifiers $\{h_{t,\lambda}\}$. The *multi-scale classifier* (F) is a combination of the set of weak classifiers $\mathcal{S}_\lambda(p)$ selected for each scale λ :

$$F(p) = \text{sign}\left(\sum_{\lambda_i} \mathcal{S}_{\lambda_i}(p)\right) = \text{sign}\left(\sum_{\lambda_i} \sum_t^T \alpha_{t,\lambda_i} h_{t,\lambda_i}(p)\right) \quad (1)$$

The strategy of building a multi-scale classifier consists of keeping a set of weights over the training set. These weights can be interpreted as a measure of the level of difficulty to classify each training sample. At the beginning, the pixels have the same weight, then in each round, the weights of misclassified pixels are increased. Thus, in the next rounds the weak learners focus on difficult samples. We will note $W_t(p)$ the weight of pixel p in round t , and $D_{t,\lambda}(R)$ the misclassification rate of region R in round t at scale λ given by the mean of the weights of its pixels:

$$D_{t,\lambda}(R) = \left(\frac{1}{|R|} \sum_{p \in R} W_t(p)\right) \quad (2)$$

Algorithm 2 presents the boosted-based training used in each stage described in Figure 3. Let $Y_\lambda(R)$, the set of labels of regions R at scale λ , be the training set. In a series of rounds $t = 1, \dots, T$, for scale λ , the weight of each region $D_{t,\lambda}(R)$ is computed (line 3). This piece of information is used to select the regions to be used for training the weak learners, building a subset of labeled regions $\hat{Y}_{t,\lambda}$ (line 5). The subset $\hat{Y}_{t,\lambda}$ is used to train the weak learners with each feature \mathcal{F} at scale λ (line 6). Each weak learner produces a weak classifier $h_{t,(\mathcal{F},\lambda)}$ (line 8). The algorithm then selects the weak classifier h_t that decreases the error $Err(h, W)$ the most (line 10). The level of error of h_t is used to compute the coefficient α_t , which indicates the degree of importance of h_t in the final classifier (line 11). The selected weak classifier h_t and the coefficient α_t are used to update the weights of the pixels $W_{(t+1)}(p)$ which can be applied in the next round (line 12).

The classification error of classifier h is:

$$Err(h, W) = \sum_{p|h(p)Y_0(p)<0} W(p) \quad (3)$$

The training is performed on the training set labels Y_λ corresponding to the same scale λ . The weak learners (linear SVM, for example) use the subset $\hat{Y}_{t,\lambda}$ for training and produce a weak classifier $h_{t,(\mathcal{F},\lambda)}$.

Algorithm 2 The boosted-based training.

Given:

Training labels $Y_\lambda(R)$ = labels of some regions R at scale λ

Initialize:

For all pixels p , $W_1(p) \leftarrow \frac{1}{|Y_0|}$, where $|Y_0|$ is the number of pixels in the image level

```

1 For  $t \leftarrow 1$  to  $T$  do
2   For all  $R \in P_\lambda$  do
3     Compute  $D_{t,\lambda}(R)$ 
4   End for
5   Build  $\hat{Y}_{t,\lambda}$  (a training subset based on  $D_{t,\lambda}(R)$ )
6   For each feature type  $\mathcal{F}$  do
7     Train weak learners using features  $(\mathcal{F}, \lambda)$  and training set  $\hat{Y}_{t,\lambda}$ .
8     Evaluate resulting classifier  $h_{t,(\mathcal{F},\lambda)}$ : compute  $Err(h_{t,(\mathcal{F},\lambda)}, W)$  (Equation 3)
9   End for
10  Select the weak classifier  $h_t$  whose  $Err = \operatorname{argmin}_{h_{t,(\mathcal{F},\lambda)}} Err(h_{t,(\mathcal{F},\lambda)}, W_{t,\lambda})$ 
11  Compute  $\alpha_t \leftarrow \frac{1}{2} \ln \left( \frac{1+r_t}{1-r_t} \right)$  with  $r_t \leftarrow \sum_p cY_0(p)h_t(p)$ 
12  Update  $W_{t+1}(p) \leftarrow \frac{W_t(p) \exp(-\alpha_t Y_0(p)h_t(p))}{\sum_p W_t(p) \exp(-\alpha_t Y_0(p)h_t(p))}$ 
13 End for

```

Output: Classifier $\mathcal{S}_\lambda(p)$

The training set labels Y_0 are the labels of pixels of image I , and training sets labels Y_λ with $\lambda > 0$ are defined according to the rate of pixels belonging to one of the two classes (for example, at least 80% of one region).

The idea of building the subset \hat{Y} is to force the classifiers to train with the most difficult samples. The weak learner should allow the most difficult samples to be differentiated from the other ones according to their weight. Thus, the strategy of creating \hat{Y} is directly dependent on the configuration of the weak classifier and may contain all regions, since the classifier considers the weights of the samples.

At the end of each stage, we withdraw the easiest samples. Let W_i be the weights of the pixels after training with scale λ_i . We denote $D_i(R_{i+1})$ the weight of region $R_{i+1} \in P_{\lambda_{i+1}}$, which is given by:

$$D_i(R_{i+1}) = \left(\frac{1}{|R|} \sum_{p \in R} W_i(p) \right) \quad (4)$$

where $W_i(p)$ is the weight of pixel $p \in R$ concerning scale λ_i .

The set of regions \check{Y}_{i+1} to be used in the training stage with scale λ_{i+1} is composed by the regions

$R_{i+1} \in P_{\lambda_{i+1}}$ with mean $D_i(R_{i+1}) > \frac{1}{2|Y_0|}$. This means that the regions that ended a training stage with distribution equal to half the initialization value $\frac{1}{|Y_0|}$ are discarded from one stage to another in the hierarchical training (Figure 3).

B. Weak Learner

We use as weak learner an SVM-based on a specific feature type \mathcal{F} and a specific scale λ . Given the training subset labels \hat{Y}_λ , the strategy is to find the best linear hyperplane of separation between the samples according to their classes (relevant and non-relevant region), trying to maximize the data separation margin. These samples are called support vectors and are found during the training. Once the support vectors and the decision coefficients $(\alpha_i, i = 1, \dots, N)$ are found, the SVM weak classifier is defined as:

$$SVM_{(\mathcal{F}, \lambda)}(R) = \text{sign}\left(\sum_i^N y_i \alpha_i (f_R \cdot f_i) + b\right) \quad (5)$$

where b is a parameter found during the training. The support vectors are the f_i such that $\alpha_i > 0$, y_i is the label of regions R_i (+1 for relevant, -1 else) and f_R is the feature vector of region R .

The training subset $\hat{Y}_{t,\lambda}$ is composed by n labels from Y_λ with values of $D_{t,\lambda}(R)$ larger or equal to $\frac{1}{|Y_0|}$. This strategy guarantees that only the regions considered as the most difficult ones are used for the training. For the first round of boosting, the regions which compose the subset $\hat{Y}_{0,\lambda}$ are randomly selected.

Given that the boosting theory requires weak classifiers (just better than random), we used a linear SVM with small subsets from the training set in each round. This strategy not only ensures the weakness of base classifiers, but also avoids the increasing of the learning time when the training set becomes larger. Moreover, in our experiments, the dimension of the feature space is smaller than the number of samples, which theoretically guarantees the weakness of linear classifiers.

C. Active Learning

Active learning is a machine learning approach which aims at obtaining high classification accuracy using very few training samples [50]. It attempts to overcome the training sample selection by asking queries in the form of unlabeled instances to be labeled by the user. The main challenge is to find the most “informative” samples, i.e., once added to the training set, the ones which lead the system to build the best classification function.

Active learning is widely used in the literature, even in remote sensing community, in applications based on SVM [28]. These approaches exploit the notion of minimum marginal hyperplane in SVMs, to select representative samples. The general strategy consists in selecting the unlabeled samples that are closer to the separation margin.

Nevertheless, many approaches have been proposed to perform active learning in boosting-based methods [49, 51–53]. We adopted the active learning strategy (active AdaBoost) proposed by Lee et al. [49]. They proposed a geometrical representation of AdaBoost output. In this representation, each sample is a point in a version space. Each point is based on the label provided by each weak learner. Therefore, each weak classifier corresponds to a dimension in this space.

Let \mathcal{S} be the output of Algorithm 2. $\mathcal{S} = 0$ can be interpreted as a separating hyperplane in the version space. The strategy proposed by Lee et al. consists in maximizing the distance of the samples to the separating hyperplane by selecting the most uncertain samples in each feedback iteration. We adapt this idea to our problem, by computing for each scale λ , the closest region to the hyperplane.

Let $\mathcal{S}_\lambda(p)$ be the output of training at scale λ , the distance of pixel p to separating hyperplane $g(p)$ is:

$$g(p) = \left| \sum_{\lambda_i} \mathcal{S}_{\lambda_i}(p) \right| = \left| \sum_{\lambda_i} \sum_t^T \alpha_{t,\lambda_i} h_{t,\lambda_i}(p) \right| \quad (6)$$

The distance of region $R \in P_\lambda$ to the separating hyperplane $g(R)$ is given by:

$$g(R) = \left(\frac{1}{|R|} \sum_{p \in R} g(p) \right) \quad (7)$$

Thus, the region corresponding to the minimal distance to the separating hyperplane g_λ^- for scale λ is defined as:

$$g_\lambda^- = \operatorname{argmin}_{R \in P_\lambda} g(R) \quad (8)$$

Equation 7 gives a measure of the degree of doubt to classify an unlabeled region. Figure 4 shows an example of classification with different classification levels. In this figure, the white regions represent the class of interest (coffee), while the black represents non-interest regions. The redder the region, the closer to the decision function, i.e., the more interesting for user feedback.

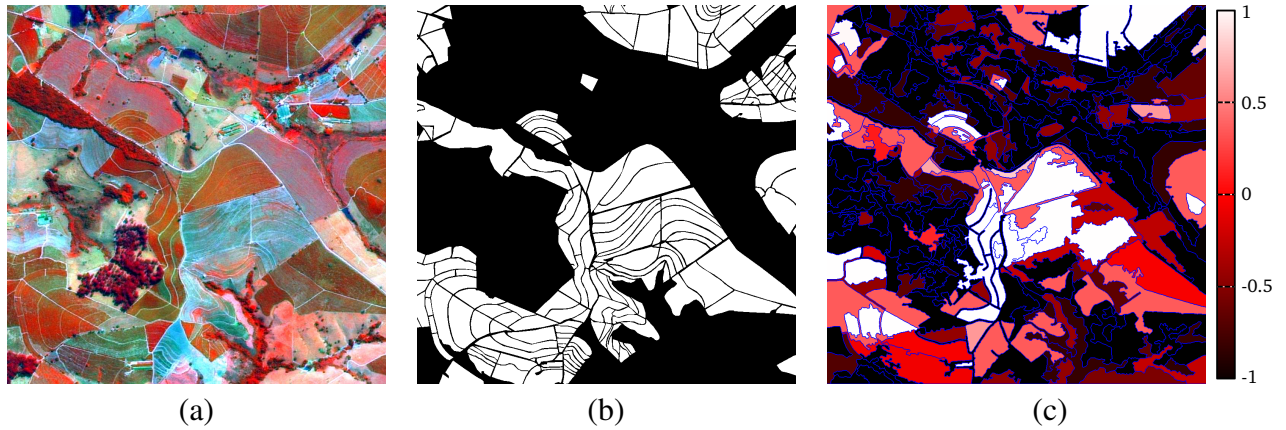


Fig. 4. Example of classification with different degrees of doubt: (a) original image, (b) ground truth and, (c) regions with different classification levels. In (b) and (c), white and black regions are coffee and non-coffee crops, respectively. The redder the region, the closer to the decision function.

D. User Interaction

Our system is strongly interactive. This means that the user is in control of the classification by introducing new examples and counter-examples to the supervised classifier at each feedback step. The classification is performed on regions of various scales, but the final result is a classification of pixels.

At first, the user has to indicate a few areas of each class. Let us remind that we have two classes, one is the class of interest (i.e, coffee) and the other one is the rest of the image (non-coffee). There are different alternatives to label pixels, from which the system is going to obtain the first region samples. The most naïve way is to label pixels as belonging to one class or to the other. It is surely a laborious and time-consuming strategy to get enough region samples to start the classification. However, this strategy can be used at the end of the classification process to refine the final classification. Another commonly used approach is to draw rectangles or polygons on the image, whose class is known for sure (examples and counter-examples). Another tool often provided to users is a *brush*, with which users can identify the target classes by painting regions on a RSI.

For all cases, the system has to translate the sets of pixels labeled by the user into a set of regions. This can be achieved by a majority vote scheme: if a region is covered by a certain percentage (for example more than 80% for the coarse scale) of pixels indicated by the user as belonging to one class, the region is used as example of this class.

Of course when the image is segmented, it is faster to directly annotate regions as examples or counter-examples for the current query [54]. In the simulation of interaction we present in the experiment section, we cannot use the regions, since our system works with several scales of segmentation. Therefore, we used

rectangles drawn inside regions whose label is known for sure. During the feedback iterations, intermediate results of classification are displayed to the user. The method selects a region at each scale. The number of regions may be lower than the number of scales if there is an intersection between the selected regions at two or more scales. In these cases, the coarsest region is selected. In our approach, the user annotates requested regions by scratching/brushing the pixels of each class as illustrated in Figure 5.

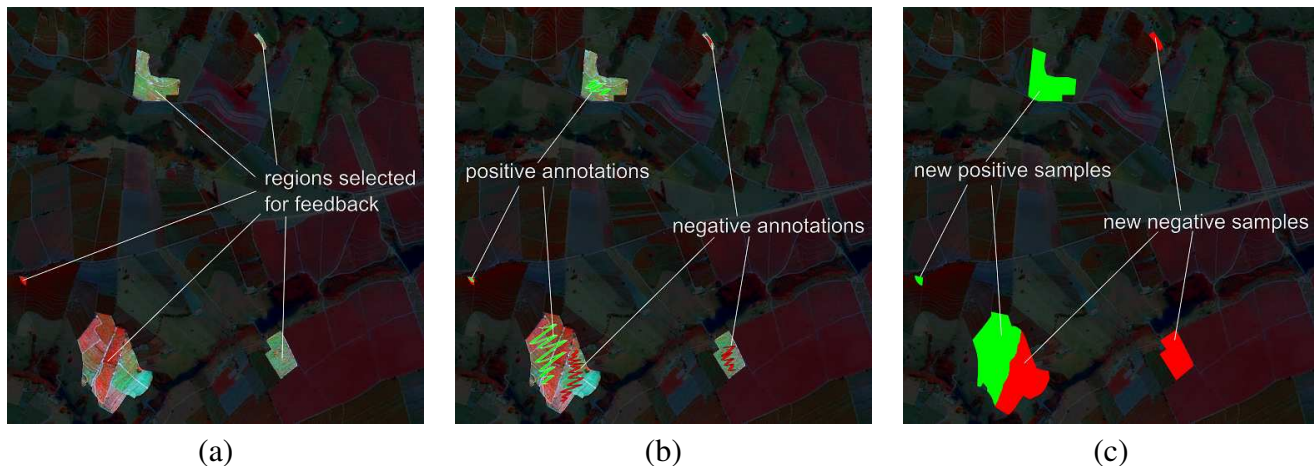


Fig. 5. Example of the process of regions annotation for user feedback: (a) regions selected for annotation; (b) user annotations, and (c) annotations converted into labelled pixels.

Figure 5 (a) illustrates the regions selected to be annotated. The user annotates the pixel classes by scratching/brushing the regions. Figure 5 (b) shows an example of annotation. In this example, positive samples are in green and negative samples are in red. The labels are then propagated to the other pixels of the selected regions as in Figure 5 (c). The remaining region pixels receive the same label of the nearest pixel annotated by the user.

V. EXPERIMENTS

In this section, we present the experiments performed to validate our method. They were carried out to address the following research questions:

- Is the proposed multi-scale approach for interactive classification effective in RSI classification tasks (Section V-B2)?
- Is the interactive method more effective than supervised classifiers built on a large training set (Section V-B3)?

In Section V-A, we describe the basic configuration of our experiments. In Section V-B we present the experimental results.

TABLE I
REMOTE SENSING DATASETS USED IN THE EXPERIMENTS.

	COFFEE	URBAN
Terrain	mountainous	plain
Satellite	SPOT	QuickBird
Spatial resolution	2.5m	0.6m
Bands composition	NIR-R-G	R-G-B
Acquisition date	08–29–2005	2003
Location	Monte Santo county, MG	Campinas,SP

A. Setup

1) *Dataset*: We have used two different datasets in our experiments. We refer to the dataset according to the target regions: COFFEE and URBAN. Table I presents a brief overview of each image. The datasets are described in details in the following sections.

a) *COFFEE dataset*: This dataset is a composition of scenes taken by the SPOT sensor in 2005 over Monte Santo de Minas county, State of Minas Gerais, Brazil. This area is a traditional place of coffee cultivation, characterized by its mountainous terrain. In addition to common issues in the area of pattern recognition in remote sensing images, these factors add further problems that must be taken into account. In mountainous areas, spectral patterns tend to be affected by the topographical differences and by interferences generated by shadows. This dataset provides an ideal environment for multi-scale analysis, since the variations in topography require the cultivation of coffee in different crop sizes. Another problem is that coffee is not an annual crop. This means that, in the same area, there are plantations of different ages. In terms of classification, we have several completely different patterns representing the same class while some of these patterns are much closer to other classes.

We have used a complete mapping of the coffee areas in the dataset to simulate the user in the experiments. The identification of coffee crops was manually done in the whole county by agricultural researchers. They used the original image as reference and visited the place to compose the final result.

The dataset is composed of 5 subimages that cover the studied region. Each image is composed by 1 million pixels (1000×1000) with spatial resolution equal to 2.5 meters.

The results of the experiments described in Section V-B2 were obtained considering all combinations of the five images used, training with three of them and testing in the same three images. The results of the experiments described in Section V-B3 were obtained considering all combinations of the 5 used subimages, training with 3 of them and testing with 2 subimages.

b) URBAN dataset: This dataset is a Quickbird scene taken in 2003 from Campinas region, State of São Paulo, Brazil. It is composed by three bands that correspond to the visible spectrum (red, green, and blue). We have empirically created the ground truth based on our knowledge about the region. We considered as urban the places which correspond to residential, commercial or industrial regions. Highways, roads, native vegetation, crops and rural buildings are considered as non-urban areas.

In the experiments, we used 5 subimages from this region. Each image is composed by 1 million pixels (1000×1000) with spatial resolution equal to 0.62 meters. The experimental protocol is the same of the COFFEE dataset.

2) Feature Extraction: Unlike ordinary images, RSI bands do not usually correspond exactly to the human visible spectrum. To apply conventional image descriptors (which generally use three color channels), it is necessary to select the most informative bands. Therefore, for the COFFEE dataset, we used only the bands corresponding to “red”, “infrared”, and “green” that are fundamental to describe vegetation targets. These spectral bands are the basis for the most vegetation indexes [55].

We extracted seven different features from each dataset by using four color descriptors (ACC [56], BIC [57], CCV [58], and GCH [59]) and three texture descriptors (Unser [60], SID [61], and QCCH [62]).

We considered five different scales to extract features from λ_1 (the finest scale) to λ_5 (the coarsest one). We selected the scales according to the principle of dichotomic cuts (see Section III). For the COFFEE dataset, at λ_5 scale, subimages contain between 200 and 400 regions while, at scale λ_1 , they contain between 9,000 and 12,000 regions. For the URBAN dataset, at λ_5 scale, subimages contain between 40 and 100 regions while, at scale λ_1 , they contain between 4,000 and 5,000 regions. Figure 6 illustrates the multi-scale segmentation for one of the subimages from URBAN dataset.

3) Simulated user annotation: In the experiments, the ground truth for unlabeled regions are used to simulate the user annotations. A similar strategy was adopted in [7, 30], as well as in content-based image retrieval methods based on relevance feedback [37]. The initial annotation was simulated by randomly selecting a small set of contiguous pixels from the training set. In the remaining steps, we used all pixels in the selected regions as user annotations, which is the process described in Section IV-D.

4) Assessment of results: To analyze the results, we computed the overall accuracy and Kappa index for the classified images at each iteration. In our experiments, the overall accuracy is defined as the sum of true positive and true negative samples divided by the total number of samples. Kappa is an effective index to compare classified images, commonly used in RSI classification [63]. Experiments in different areas

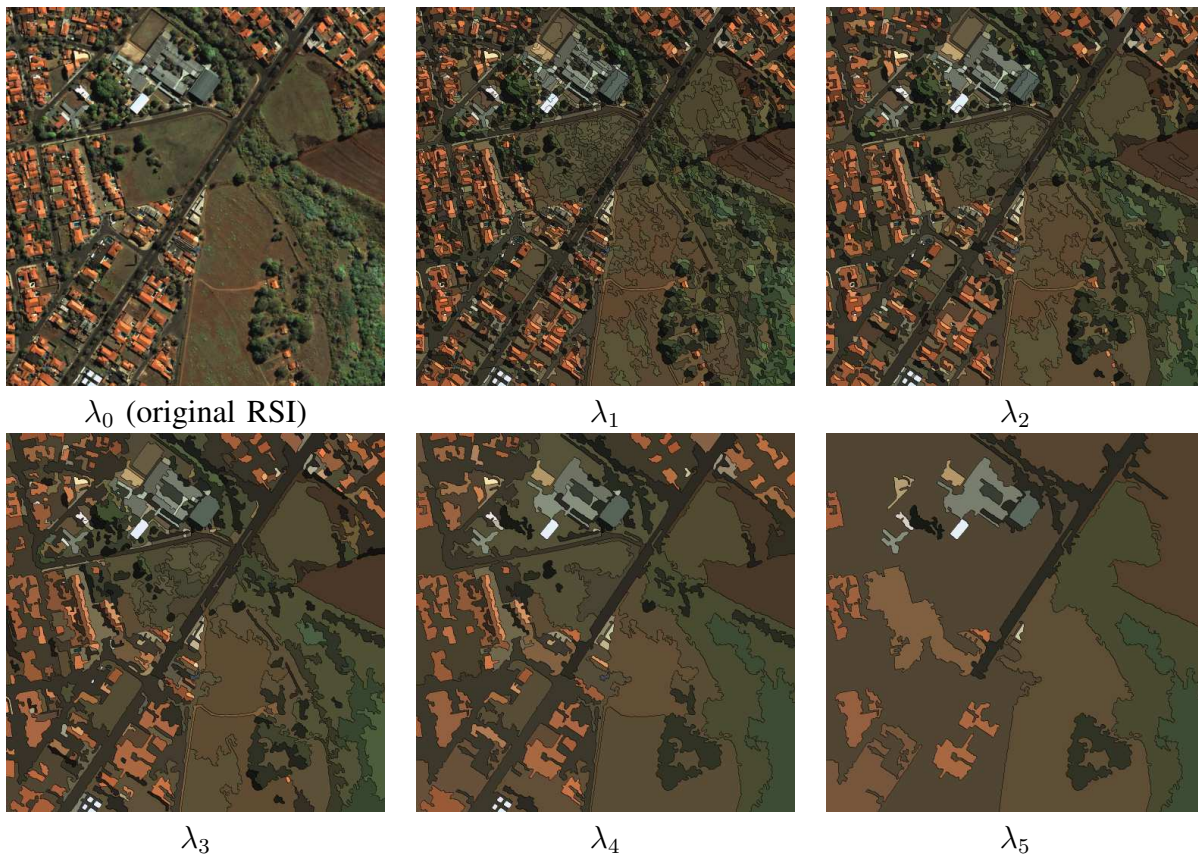


Fig. 6. One of the tested subimages and the segmentation results in each of the selected scales for the URBAN dataset.

show that Kappa could have various interpretations and these guidelines could be different depending on the application. However, Landis and Koch [64] characterize Kappa values over 0.80 as “almost perfect agreement”, 0.60 to 0.79 as “substantial agreement”, 0.40 to 0.59 as “moderate agreement”, and below 0.40 as “poor agreement”. Negative Kappa means that there is no agreement between classified data and verification data.

B. Results

1) *Interactive Classification Example:* In this section, we present an example of a result of the proposed method for interactive classification. Figure 7 presents the results for one of the tested images from the COFFEE dataset compared to the original image and the ground truth. This image is composed of several regions of coffee, pasture, native forest, and some lakes.

As the method begins with a very small training set, the “Initial Result” is visually different from the ground truth. One reason is that the training set may not have been large enough to correctly classify regions. With the gradual increase in the training set, the results improve until the fourth iteration (OA=83.55% κ =0.8031). Between the fifth and the ninth feedback steps, we can note many variations

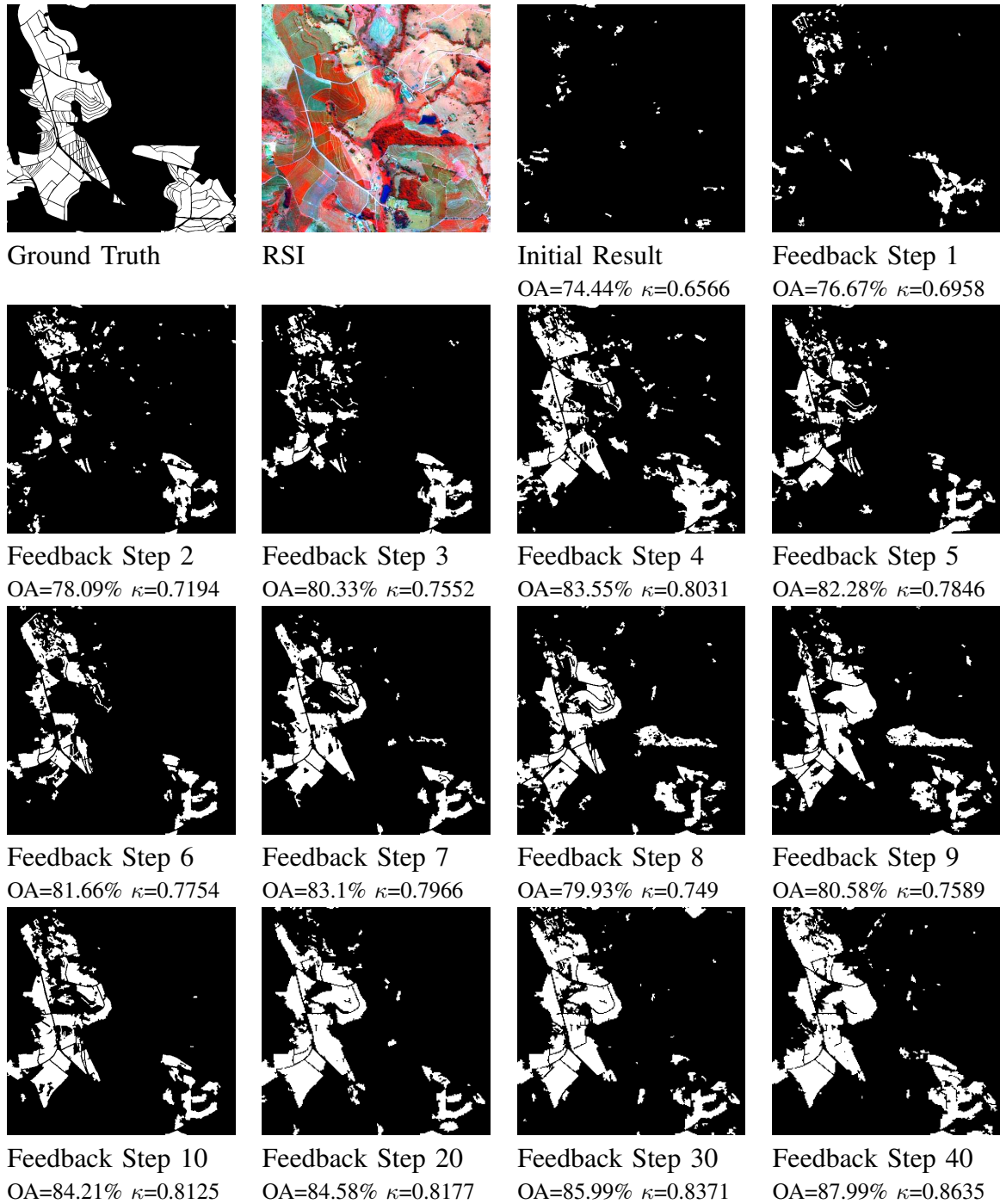


Fig. 7. Example of the results from the initial classification to the feedback step 10, 20, 30, and 40 compared to the original image and the ground truth. Coffee and non-coffee regions are represented in white and in black respectively. OA=Overall Accuracy; κ =Kappa index.

in the results due to confusion between: 1) “new coffee” crops and pasture; and 2) “mature coffee” and native forest. The result is improved and becomes more stable from the tenth feedback step on. Although the improvements are smaller, they continue along the iterations, as it can be seen from the results of feedback steps 20, 30, 40 and so on.

To better illustrate the results, Figure 8 presents an error analysis (false positive and false negative samples) for the result in feedback steps 9, 10, and 40. Table II presents the accuracy values.

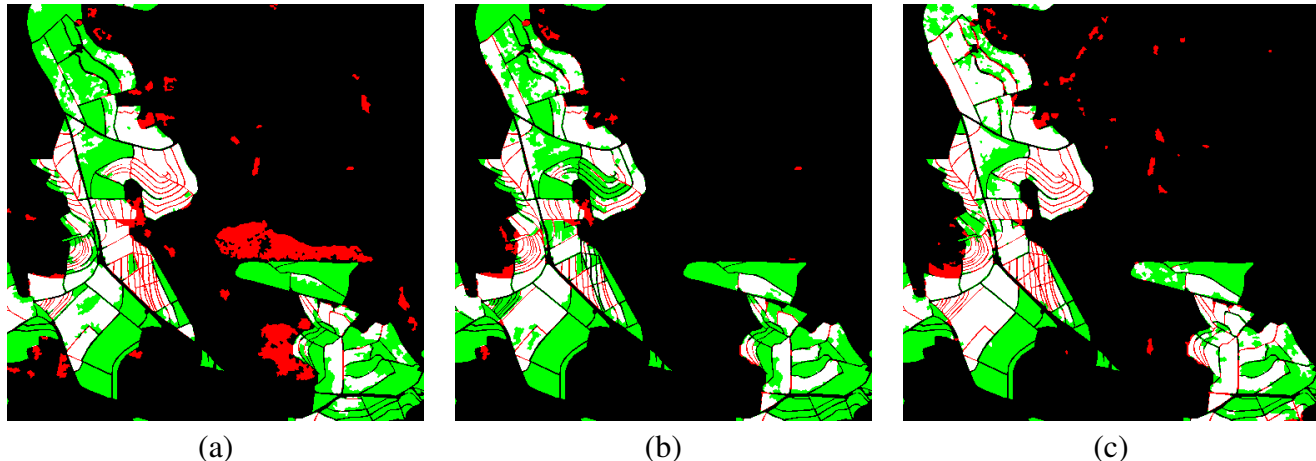


Fig. 8. A result obtained with the proposed method in feedback steps 9 (a), 10 (b), and 40 (c) for the experiment presented in Figure 7. Pixels correctly classified are shown in white (true positive) and black (true negative) while the errors are displayed in red (false positive) and green (false negative).

TABLE II

ACCURACY ANALYSIS OF CLASSIFICATION FOR THE EXAMPLE PRESENTED IN FIGURE 8 (TP = TRUE POSITIVE, TN = TRUE NEGATIVE, FP = FALSE POSITIVE, FN = FALSE NEGATIVE).

Feedback Step	TP	TN	FP	FN
9	117,322	688,431	51,737	142,510
10	119,225	722,912	17,256	140,607
40	167,255	712,611	27,557	92,577

From the feedback steps 9 to 10, one can notice a great reduction in the number of false positives (red pixels). Most of the removed pixels correspond to areas of natural vegetation. This indicates that the confusion between natural vegetation and mature coffee is reduced. Comparing Figure 8 (b) and Figure 8 (c), we note that the classification seems to go through a refining process. Visually, it is possible to see small difference between the results in feedback steps 10 and 40. However, in Table II, we can observe that the number of pixels corresponding to coffee regions significantly increased (from 119,225 to 167,255).

As far as time is concerned, experiments with the COFFEE dataset showed that the proposed method takes around 50s for each training step using one scale and the combination of the seven weak classifiers. The proposed method needs less than one hour to perform 10 steps using five scales. Considering that 10 feedback steps is a good number to get a satisfactory result of classification, one hour is not much if compared with the time usually spent to perform manual mapping of large areas. Furthermore, the steps to evaluate each descriptor in the method is easily parallelizable and, hence, the training time in each

interaction can be reduced in a real scenario.

Regarding the URBAN dataset, the IHMSC needs 12s to train on each scale since it has less regions.

2) *Multi-Scale versus Individual Scale*: In this section, we compare the classification results obtained by using individual scales against the combination of scales by using the IHMSC approach presented in Section IV-A with 20 rounds for each scale. We used IHMSC to perform the individual scales experiment with 100 rounds. In these experiments, we tested all descriptors referenced in Section V-A2. The initial training set is a rectangle composed by 10,000 pixels with both classes.

Figure 9 presents the Kappa \times Feedback Steps curves for the COFFEE dataset. Figure 10 shows the Overall Accuracy \times Feedback Steps curves for the COFFEE dataset.

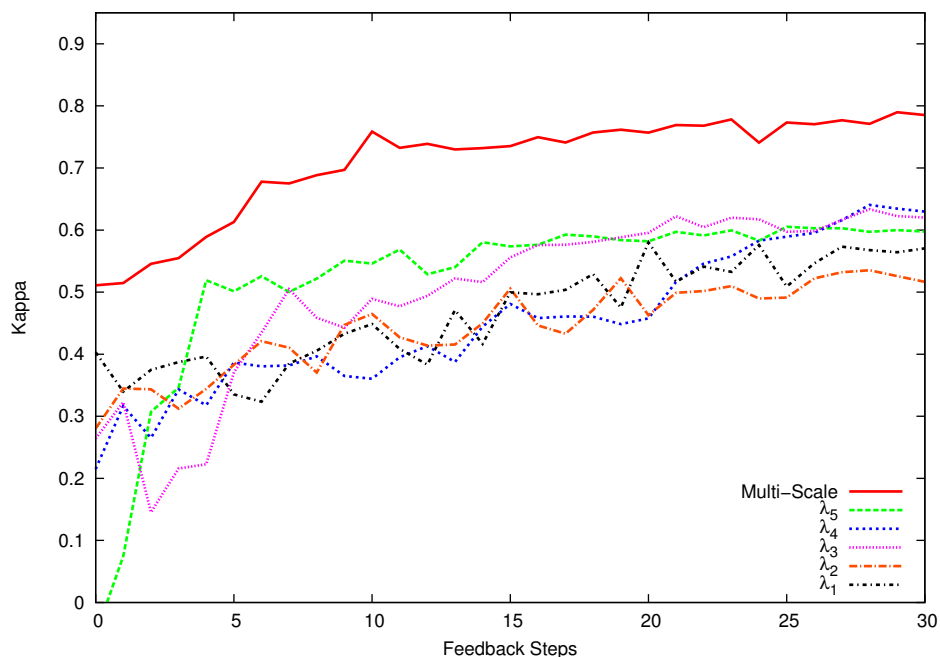


Fig. 9. Kappa index for each iteration of feedback for the COFFEE dataset considering five scales and the multi-scale classification approach.

According to the results for the COFFEE dataset, one can observe that the combination of scales presents better results than individual ones. We can note that intermediate scales (λ_4 , λ_3) use more iterations to converge, but achieve better results after many feedback steps. Concerning the coarser scale (λ_5), it quickly obtains good results, but there is no improvement after 14 feedback steps. In this scenario, regions of interest for training in the coarse scales are more quickly exhausted. We conclude that the IHMSC method yields reasonable results with few feedback steps. It is even able to improve them later as it allows the refinement of the training set along iterations.

Figures 11 and 12 present the Kappa \times Feedback Steps curves and Overall Accuracy \times Feedback Steps

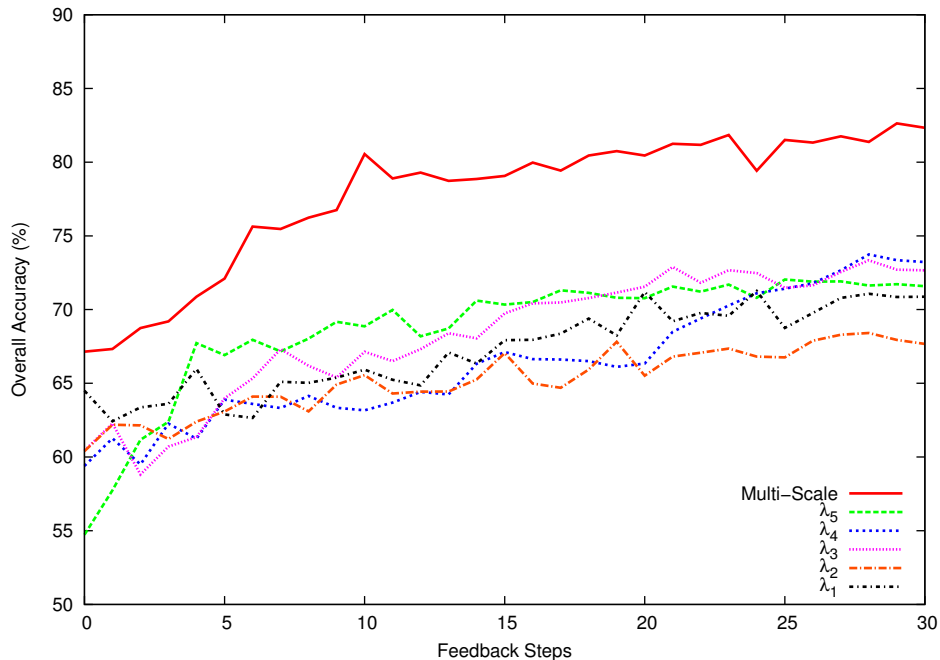


Fig. 10. Overall accuracy for each iteration of feedback for the COFFEE dataset, considering five scales and the multi-scale classification approach.

curves, respectively, for the URBAN dataset.

We can observe that, for the URBAN dataset, multiscale training achieves results that are better than the ones for individual scales, except for the two first feedback steps in which the training set is too small. Coarse scales produce better results than finer ones. Which means that the features extracted from finer scales can not properly represent the urban areas. However, it is not difficult to understand this phenomenon. Urban areas are complex targets since they are composed by smaller objects with specific characteristics. If we use a fine scale, small objects (e.g., trees) can be present in both urban and non-urban areas. This fact makes the classification task more difficult.

3) Interactive versus Supervised Classification Strategy: In this section, we present experiments that compare the proposed method for interactive classification with a traditional supervised approach that uses the whole available training set. For this reason, the experiments of this section (including the interactive method) were performed using all combinations of the five images from our dataset: three images as training set and two images for testing. The difference is that the supervised method uses all the available training images to learn while the same information is used to simulate the user annotations in the interactive approach.

It is important to note that in a real situation the user would typically annotate and classify regions

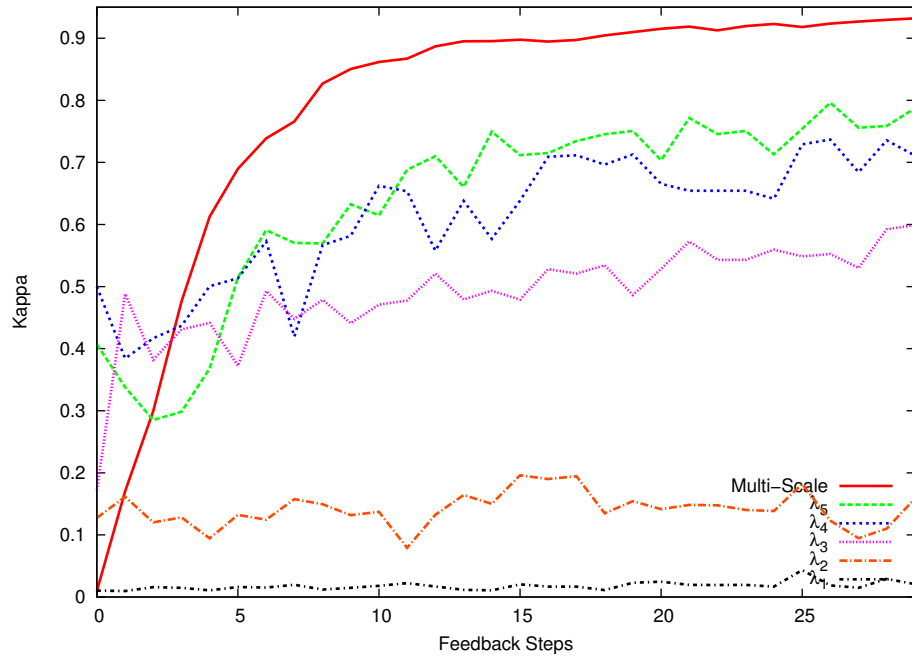


Fig. 11. Kappa index for each iteration of feedback for the URBAN dataset, considering five scales and the multi-scale classification approach.

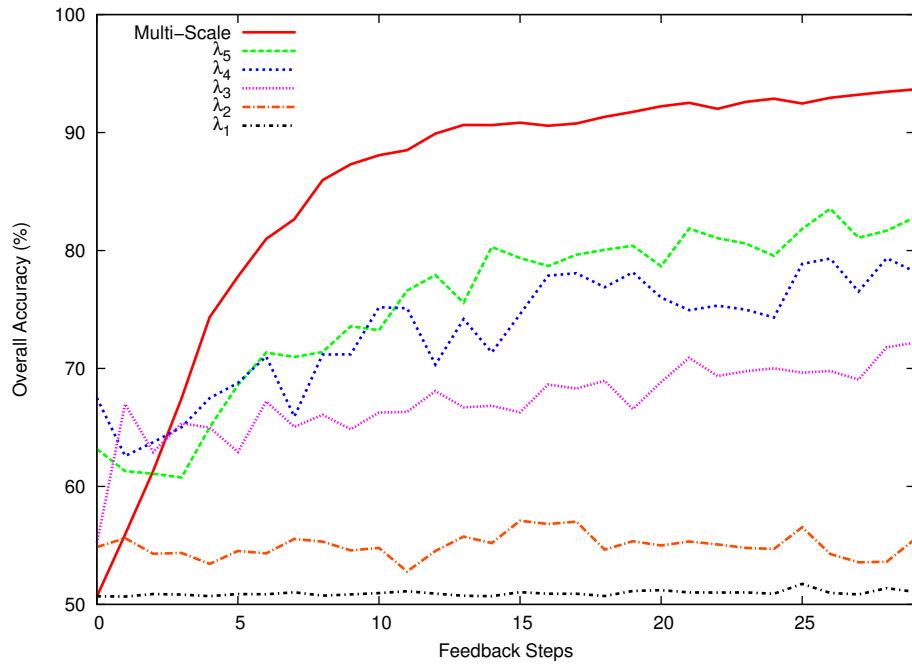


Fig. 12. Overall accuracy for each iteration of feedback for URBAN dataset, considering five scales and the multi-scale classification approach.

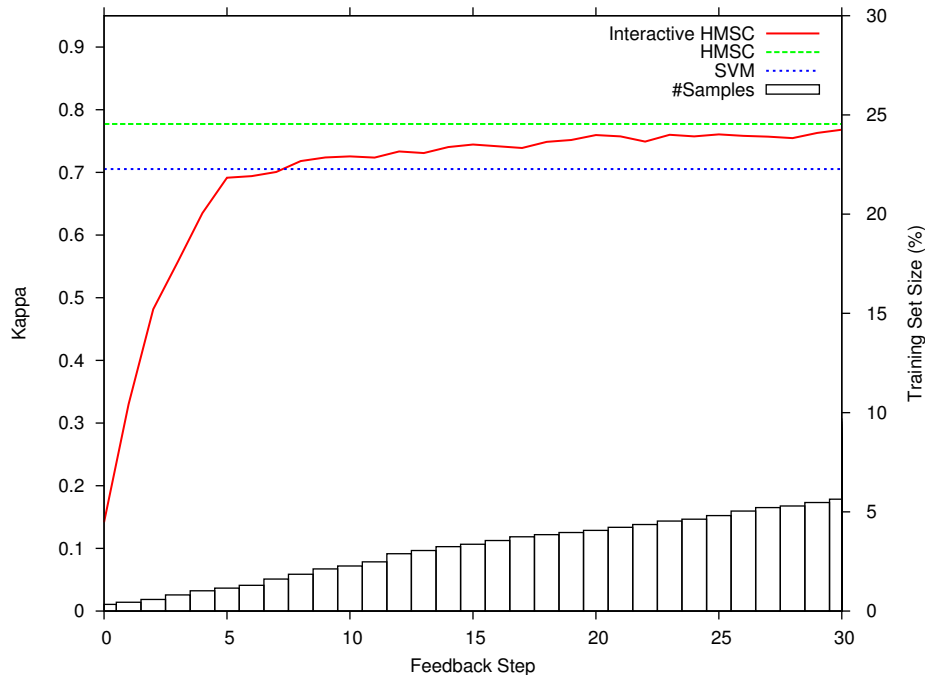


Fig. 13. Kappa index for the HMSC and SVM and Kappa \times Feedback Steps curves for interactive HMSC using the COFFEE dataset. The histogram represents the percentage of the training set used in the interactive method.

present in the same image scenes like in the experiments reported in Section V-B2.

We used two supervised methods as baselines. The first method is the HMSC with no user interactions using 100% of the pixels available for training (3,000,000 of pixels in this experiments). The other method is based on Tzotsos et al. [65]. They proposed a method that uses SVMs with RBF kernels to classify the regions obtained from a multi-scale segmentation process. That approach outperforms the results obtained by using the software eCognition [4]. Therefore, we used SVM + RBF kernels applied to an intermediate segmentation scale defined by the Guigues method as baseline. The BIC descriptor was used in this baseline.

Figure 13 presents the classification results for the baselines and the Kappa \times Feedback Steps curves using the COFFEE dataset. This figure also includes the histogram of the percentage of the training set used in each feedback step by the proposed interactive method. Figure 14 presents the same classification results using the COFFEE dataset, but using the Overall Accuracy. The interactive HMSC training set starts with two rectangles composed of 5,000 pixels for each class (coffee and non-coffee). It corresponds to 0.33% of the training set.

According to the results of Figures 13 and 14, HMSC has Kappa equal to 0.77 and overall accuracy equal to 82%. SVM has Kappa equal to 0.71 and overall accuracy equal to 77%. The interactive method

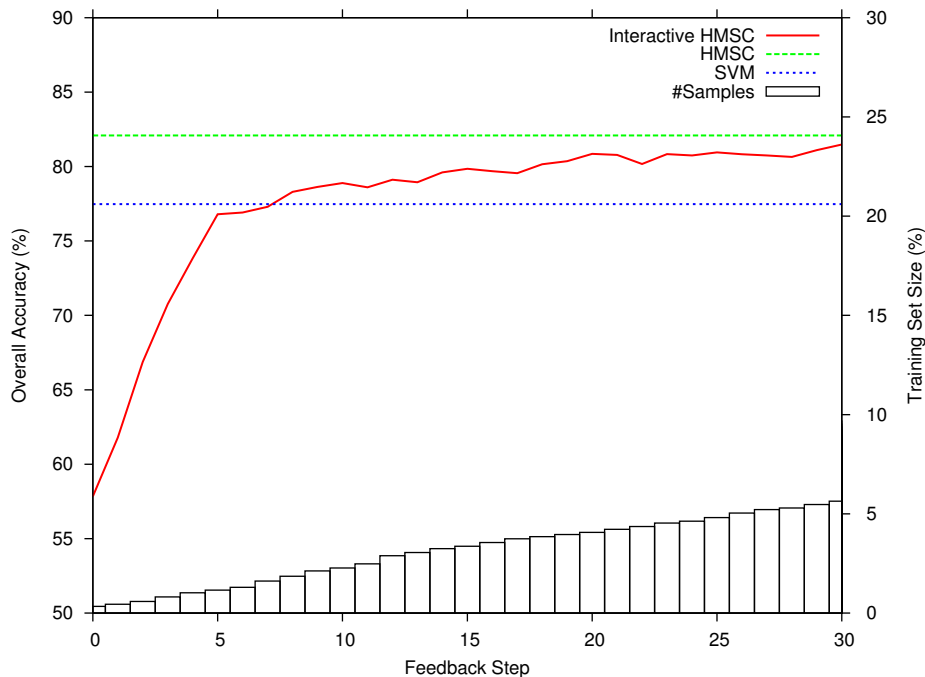


Fig. 14. Overall Accuracy results for the HMSC and SVM and Overall Accuracy \times Feedback Steps curves for interactive HMSC using the COFFEE dataset. The histogram represents the percentage of the training set used in the interactive method.

starts with Kappa index equal to 0.15 and overall accuracy equal to 57%. After 20 iterations, the results converge to Kappa index equal to 0.76 and overall accuracy equal to 81%.

One can note that the interactive *HMSC* obtains similar results to the SVM baseline after 5 feedback steps. After 20 feedback steps, interactive HMSC obtains results close to the supervised HMSC. Therefore, these experiments show that by using about 1% of the pixels in the training set, we can obtain results close to SVM. By using a little bit more than 5% of the training set, the interactive method can achieve the same results as the HMSC trained with the whole set.

Figure 15 presents the classification results for the baselines and the Kappa \times Feedback Steps curves using the URBAN dataset. Figure 16 presents the same classification results using the URBAN dataset, considering the Overall Accuracy.

In general, the conclusions for the URBAN dataset are similar to the results obtained for the COFFEE dataset. With few iterations, the IHMSC achieves classification results as good as the SVM baseline. With some more feedback steps, by using 15% of the training set, the interactive approach achieved almost the same accuracy (88%) obtained by the HMSC supervised approach.

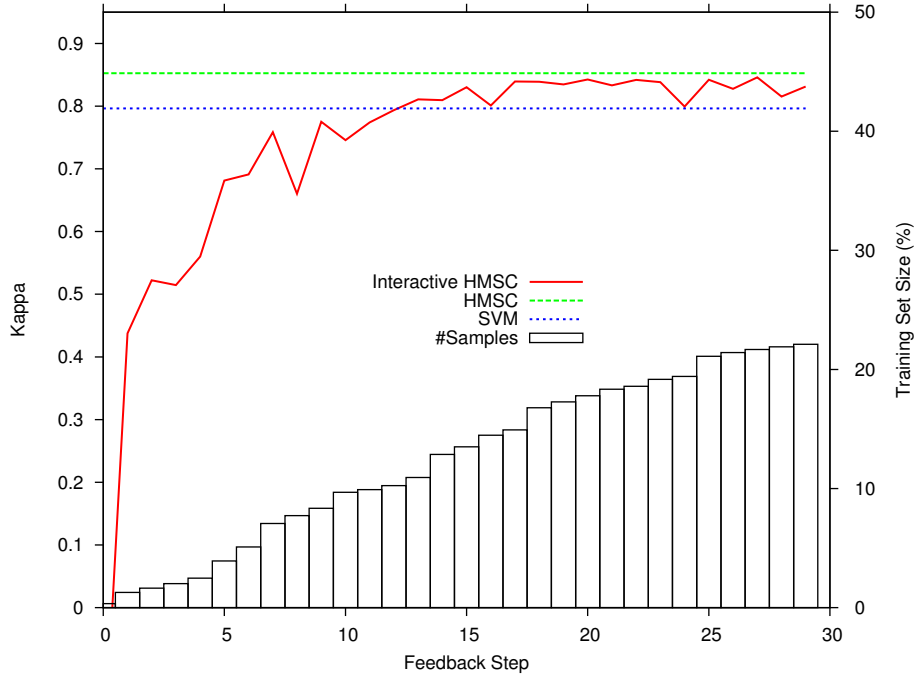


Fig. 15. Kappa index for the HMSC and SVM and Kappa \times Feedback Steps curves for interactive HMSC using the URBAN dataset. The histogram represents the percentage of the training set used in the interactive method.

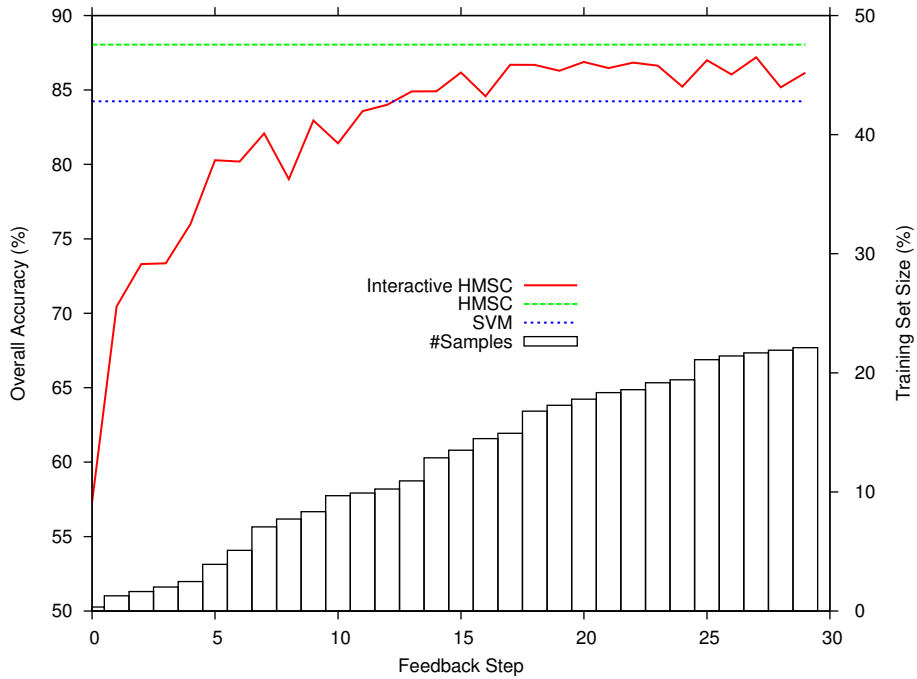


Fig. 16. Overall Accuracy results for the HMSC and SVM and Overall Accuracy \times Feedback Steps curves for interactive HMSC using the URBAN dataset. The histogram represents the percentage of the training set used in the interactive method.

VI. CONCLUSIONS

We have shown in this paper that interactive classification based on active learning can be a good alternative to the selection of a suitable training set for high resolution remote sensing analysis. We also showed that by combining multi-scale features one can achieve high classification accuracy.

We proposed a method for interactive classification of remote sensing images considering multi-scale segmentation: the interactive HMSC (Hierarchical Multi-Scale Classifier). The objective is to improve the selection of training samples by using the features from the most appropriate scales of representation.

HMSC uses a hierarchical region segmentation from the pixel level to the image level to perform multi-scale segmentation. The training strategy is based on the AdaBoost algorithm, which builds a linear combination of weak classifiers (in this work, we use linear SVMs). Each classifier is trained for each feature vector extracted from regions at each scale of segmentation. During a feedback step, for each considered scale, the method selects the regions that are closer to the separating border.

The experiments showed that the combination of scales produce better results than isolated scales in a relevance feedback process. The interactive HMSC achieves more than 80% of accuracy with 10 iterations in both used datasets, overcoming the baseline based on SVM. By using a little bit more than 5% of the training set for the COFFEE dataset and 10% for the URBAN dataset, the interactive method can achieve the same results as the supervised HMSC trained with the whole set.

Future works include the design and the implementation of a graphic user interface to allow experiments with a real user. We also intend to explore other strategies to select more than one region per scale in a single feedback step.

ACKNOWLEDGMENTS

The authors are grateful to CAPES/COFECUB (BEX 5224101), FAPESP (2008/58528-2), and CNPq for the financial support. We are grateful to Jean Pierre Cocquerez for the support concerning the segmentation tool. We also thank Rubens Lamparelli and Cooxupé for their support related to agricultural aspects and the remote sensing dataset.

REFERENCES

- [1] J. A. dos Santos, F. A. Faria, R. T. Calumby, R. da S. Torres, and R. A. C. Lamparelli, "A genetic programming approach for coffee crop recognition," in *IGARSS 2010*, Honolulu, USA, July 2010, pp. 3418–3421.

- [2] W. Zhou, G. Huang, A. Troy, and M. Cadenasso, "Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: A comparison study," *Remote Sensing of Environment*, vol. 113, no. 8, pp. 1769 – 1777, 2009.
- [3] R. Showengetdt, *Techniques for Image Processing and Classification in Remote Sensing*. New York: Academic Press, 1983.
- [4] 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 Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 3-4, pp. 239 – 258, 2004, integration of Geodata and Imagery for Automated Refinement and Update of Spatial Databases.
- [5] Y. Tarabalka, J. Tilton, J. Benediktsson, and J. Chanussot, "A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 1, pp. 262 –272, feb. 2012.
- [6] A. Tzotsos, K. Karantzalos, and D. Argialas, "Object-based image analysis through nonlinear scale-space filtering," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 1, pp. 2–16, 2011.
- [7] J. A. dos Santos, A. T. da Silva, R. da Silva Torres, A. X. Falcão, L. P. Magalhães, and R. A. C. Lamparelli, "Interactive classification of remote sensing images by using optimum-path forest and genetic programming," in *CAIP (2)*, 2011, pp. 300–307.
- [8] Y. O. Ouma, S. Josaphat, and R. Tateishi, "Multiscale remote sensing data segmentation and post-segmentation change detection based on logical modeling: Theoretical exposition and experimental results for forestland cover change analysis," *Computers and Geosciences*, vol. 34, no. 7, pp. 715 – 737, 2008.
- [9] G. Wilkinson, "Results and implications of a study of fifteen years of satellite image classification experiments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 433–440, March 2005.
- [10] 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, 2007.
- [11] J. Lee and T. A. Warner, "Image classification with a region based approach in high spatial resolution imagery," in *Int. Archives of Photogrammetry, Remote Sensing and Spatial Inf. Sciences*, Istanbul, Turkey, July 2004, pp. 181–187.
- [12] A. Katartzis, I. Vanhamel, and H. Sahlí, "A hierarchical markovian model for multiscale region-based classification of vector-valued images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 548–558, March 2005.
- [13] Q. Yu, P. Gong, N. Clinton, G. Biging, M. Kelly, and D. Schirokauer, "Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery," *Photogrammetric Engineering Remote Sensing*, vol. 72, no. 7, pp. 799–811, 2006.
- [14] Castillejo-González, López-Granados, García-Ferrer, Peña-Barragán, Jurado-Expósito, de la Orden, and González-Audicana, "Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using quickbird imagery," *Computers and Electronics in*, vol. 68, no. 2, pp. 207 – 215, 2009.
- [15] S. W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng, "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery," *Remote Sensing of Environment*, vol. 115, no. 5, pp. 1145 – 1161, 2011.
- [16] A. Alonso-González, S. Valero, J. Chanussot, C. López-Martínez, and P. Salembier, "Processing multidimensional sar and hyperspectral images with binary partition tree," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. PP, no. 99, pp. 1 –25, 2012.
- [17] F. Petitjean, C. Kurtz, N. Passat, and P. Ganarski, "Spatio-temporal reasoning for the classification of satellite image time series," *Pattern Recognition Letters*, vol. 33, no. 13, pp. 1805 – 1815, 2012.
- [18] J. A. dos Santos, P. Gosselin, S. Philipp-Foliguet, R. da S. Torres, and A. X. Falcão, "Multiscale classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, pp. 3764–3775, 2012.
- [19] Y. Chen, W. Su, J. Li, and Z. Sun, "Hierarchical object oriented classification using very high resolution imagery and lidar data over urban areas," *Advances in Space Research*, vol. 43, no. 7, pp. 1101 – 1110, 2009.

- [20] J. Chen, D. Pan, and Z. Mao, "Image-object detectable in multiscale analysis on high-resolution remotely sensed imagery," *International Journal of Remote Sensing*, vol. 30, no. 14, pp. 3585–3602, 2009.
- [21] M. Kim, T. A. Warner, M. Madden, and D. S. Atkinson, "Multi-scale geobia with very high spatial resolution digital aerial imagery: scale, texture and image objects," *International Journal of Remote Sensing*, vol. 32, no. 10, pp. 2825–2850, 2011.
- [22] W. jie Wang, Z. ming Zhao, and H. qing Zhu, "Object-oriented change detection method based on multi-scale and multi-feature fusion," in *Urban Remote Sensing Event, 2009 Joint*, may 2009, pp. 1–5.
- [23] T. Hermosilla, L. Ruiz, J. Recio, and M. Cambra-Lpez, "Assessing contextual descriptive features for plot-based classification of urban areas," *Landscape and Urban Planning*, 2012.
- [24] O. Rozenstein and A. Karnieli, "Comparison of methods for land-use classification incorporating remote sensing and gis inputs," *Applied Geography*, vol. 31, no. 2, pp. 533–544, 2011.
- [25] T. Novack, H. Kux, R. Feitosa, and G. Costa, "Per block urban land use interpretation using optical vhr data and the knowledge-based system interimage," in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII-4/C7*, Ghent, Belgium, 2010.
- [26] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, "A survey of active learning algorithms for supervised remote sensing image classification," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 3, pp. 606–617, june 2011.
- [27] S. Rajan, J. Ghosh, and M. Crawford, "An active learning approach to hyperspectral data classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 4, pp. 1231–1242, april 2008.
- [28] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 6, p. 2767, june 2010.
- [29] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 3, pp. 1014–1031, march 2011.
- [30] J. A. dos Santos, C. D. Ferreira, R. da S. Torres, M. A. Gonçalves, and R. A. C. Lamparelli, "A relevance feedback method based on genetic programming for classification of remote sensing images," *Information Sciences*, vol. 181, no. 13, pp. 2671–2684, 2011.
- [31] D. Tuia, E. Pasolli, and W. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sensing of Environment*, vol. 115, no. 9, pp. 2232–2242, 2011.
- [32] E. Pasolli, F. Melgani, and Y. Bazi, "Support vector machine active learning through significance space construction," *Geoscience and Remote Sensing Letters, IEEE*, vol. 8, no. 3, pp. 431–435, may 2011.
- [33] D. Stavrakoudis, J. Theocharis, and G. Zalidis, "A boosted genetic fuzzy classifier for land cover classification of remote sensing imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 4, pp. 529–544, 2011.
- [34] S. Durbha and R. King, "Semantics-enabled framework for knowledge discovery from earth observation data archives," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 11, pp. 2563–2572, nov. 2005.
- [35] V. Shah, N. Younan, S. Durbha, and R. King, "Feature identification via a combined ICA-wavelet method for image information mining," *Geoscience and Remote Sensing Letters, IEEE*, vol. 7, no. 1, pp. 18–22, jan. 2010.
- [36] J. A. dos Santos, R. A. C. Lamparelli, and R. da S. Torres, "Using relevance feedback for classifying remote sensing images," in *XIV Brazilian Remote Sensing Symposium*, Natal, RN, Brazil, Abril 2009, pp. 7909–7916.
- [37] C. D. Ferreira, J. A. dos Santos, R. da S. Torres, M. A. Gonçalves, R. C. Rezende, and W. Fan, "Relevance feedback based on genetic programming for image retrieval," *Pattern Recognition Letters*, vol. 32, no. 1, pp. 27–37, 2011.
- [38] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 1, pp. 2–16, 2010.

- [39] R. Gaetano, G. Scarpa, and G. Poggi, "Hierarchical texture-based segmentation of multiresolution remote-sensing images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 7, pp. 2129–2141, July 2009.
- [40] Z. Wang, J. R. Jensen, and J. Im, "An automatic region-based image segmentation algorithm for remote sensing applications," *Environ. Model. Softw.*, vol. 25, pp. 1149–1165, October 2010.
- [41] N. Li, H. Huo, and T. Fang, "A novel texture-preceded segmentation algorithm for high-resolution imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, no. 99, pp. 1–11, 2010.
- [42] J. Papa, A. Falcão, and C. Suzuki, "Supervised pattern classification based on optimum-path forest," *International Journal of Imaging Systems and Technology*, vol. 19, no. 2, pp. 120–131, 2009.
- [43] A. Tzotsos, C. Iosifidis, and D. Argialas, "A hybrid texture-based and region-based multi-scale image segmentation algorithm," in *Object-Based Image Analysis*, ser. Lecture Notes in Geoinformation and Cartography, T. Blaschke, S. Lang, and G. J. Hay, Eds. Springer Berlin Heidelberg, 2008, pp. 221–236.
- [44] H. Li, H. Gu, Y. Han, and J. Yang, "An efficient multiscale srmhr (statistical region merging and minimum heterogeneity rule) segmentation method for high-resolution remote sensing imagery," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 2, no. 2, pp. 67–73, June 2009.
- [45] L. Guigues, J. Cocquerez, and H. Le Men, "Scale-sets image analysis," *International Journal of Computer Vision*, vol. 68, pp. 289–317, 2006.
- [46] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on Pure and Applied Mathematics*, vol. 42, no. 5, pp. 577–685, 1989.
- [47] R. da S. Torres and A. X. Falcão, "Content-Based Image Retrieval: Theory and Applications," *Revista de Informática Teórica e Aplicada*, vol. 13, no. 2, pp. 161–185, 2006.
- [48] R. E. Schapire, "A brief introduction to boosting," in *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, ser. IJCAI '99, 1999, pp. 1401–1406.
- [49] L. W. X. Li and E. Sung, "Improving adaboost for classification on small training sample sets with active learning," in *Proceedings of Asian Conference on Computer Vision*, ser. ACCV 2004, 2004.
- [50] P. Gosselin and M. Cord, "Active learning methods for interactive image retrieval," *Image Processing, IEEE Transactions on*, vol. 17, no. 7, pp. 1200–1211, July 2008.
- [51] H.-j. Kim and J.-u. Kim, "Combining active learning and boosting for naive bayes text classifiers," in *Advances in Web-Age Information Management*, ser. Lecture Notes in Computer Science, vol. 3129, 2004, pp. 519–527.
- [52] Y. Lu, Q. Tian, and T. Huan, "Interactive boosting for image classification," in *Multimedia Content Analysis and Mining*, ser. Lecture Notes in Computer Science, vol. 4577, 2007, pp. 315–324.
- [53] J. Yu, Y. Lu, Y. Xu, N. Sebe, and Q. Tian, "Integrating relevance feedback in boosting for content-based image retrieval," in *ICASSP (1)'07*, 2007, pp. 965–968.
- [54] P. Gosselin, M. Cord, and S. Philipp-Foliguet, "Kernels on bags of fuzzy regions for fast object retrieval," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 1, 16 2007–Oct. 19 2007, pp. I–177–I–180.
- [55] T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, *Remote Sensing and Image Interpretation*. Wiley, 2007.
- [56] J. Huang, S. R. Kumar, M. Mitra, W. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 1997, p. 762.
- [57] R. de O. Stehling, M. A. Nascimento, and A. X. Falcão, "A compact and efficient image retrieval approach based on border/interior pixel classification," in *CIKM*, New York, NY, USA, 2002, pp. 102–109.
- [58] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *ACM Multimedia*, 1996, pp. 65–73.

- [59] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.
- [60] M. Unser, "Sum and difference histograms for texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 118–125, 1986.
- [61] J. A. M. Zegarra, N. J. Leite, and R. d. S. Torres, "Rotation-invariant and scale-invariant steerable pyramid decomposition for texture image retrieval," in *XX Brazilian Symposium on Computer Graphics and Image Processing*, 2007, pp. 121–128.
- [62] C. Huang and Q. Liu, "An orientation independent texture descriptor for image retrieval," *International Conference on Communications, Circuits and Systems*, pp. 772–776, July 2007.
- [63] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Washington, DC: Lewis Publishers, 1977.
- [64] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, March 1977.
- [65] A. Tzotsos and D. Argialas, "Support vector machine classification for object-based image analysis," in *Object-Based Image Analysis*, ser. Lecture Notes in Geoinformation and Cartography, T. Blaschke, S. Lang, and G. J. Hay, Eds. Springer Berlin Heidelberg, 2008, pp. 663–677.