A Novel Protocol for Accuracy Assessment in Classification of Very High Resolution Images

Claudio Persello, *Student Member, IEEE*, and Lorenzo Bruzzone, *Senior Member, IEEE*

Abstract—This paper presents a novel protocol for the accuracy assessment of the thematic maps obtained by the classification of very high resolution images. As the thematic accuracy alone is not sufficient to adequately characterize the geometrical properties of high-resolution classification maps, we propose a protocol that is based on the analysis of two families of indices: 1) the traditional thematic accuracy indices and 2) a set of novel geometric indices that model different geometric properties of the objects recognized in the map. In this context, we present a set of indices that characterize five different types of geometric errors in the classification map: 1) oversegmentation; 2) undersegmentation; 3) edge location; 4) shape distortion; and 5) fragmentation. Moreover, we propose a new approach for tuning the free parameters of supervised classifiers on the basis of a multiobjective criterion function that aims at selecting the parameter values that result in the classification map that jointly optimize thematic and geometric error indices. Experimental results obtained on QuickBird images show the effectiveness of the proposed protocol in selecting classification maps characterized by a better tradeoff between thematic and geometric accuracies than standard procedures based only on thematic accuracy measures. In addition, results obtained with support vector machine classifiers confirm the effectiveness of the proposed multiobjective technique for the selection of free-parameter values for the classification algorithm.

Index Terms—Accuracy assessment, classification maps, geometric accuracy, image classification, remote sensing, thematic accuracy, very high resolution (VHR) images.

I. INTRODUCTION

W ITH the availability of very high resolution (VHR) images acquired by satellite multispectral scanners (e.g., GeoEye-1, QuickBird, IKONOS, and SPOT 5), it is possible to acquire detailed information on the shape and geometry of the objects that are present on the ground. This detailed information can be exploited by automatic classification systems to generate land-cover maps that exhibit a high degree of geometrical details. The precision that the classification system can afford in the characterization of the geometrical properties of the objects that are present on the ground is particularly relevant in many practical applications, e.g., in urban-area mapping, building characterization, target detection, crop-field classification in precision farming, etc.

In this context, it is necessary to further develop both algorithms for characterizing the textural and geometric information

that is present in VHR images, and effective classification techniques that are capable of exploiting these properties for increasing the classification accuracy. In the literature, several techniques have been proposed for the classification of VHR images. Among others, we recall the use of texture, geometric features, and morphological transformations for characterizing the context of each single pixel, and the use of classification algorithms that can operate in large-dimensional feature spaces (e.g., support vector machine (SVM)) [1]–[5]. Nonetheless, a major open issue in classification of VHR images is the lack of adequate strategies for a precise evaluation of the quality of the produced thematic maps. The most common accuracy-assessment methodology in classification of VHR images is based on the computation of thematic accuracy measures according to the collected reference data. However, the thematic accuracy alone does not result to be sufficient for effectively characterizing the geometrical properties of the objects recognized in a map, because it assesses the correctness of the land-cover labels of sparse test pixels (or regions of interests) that do not model the actual shape of the objects in the scene. Thus, often, maps derived by different classifiers (or with different parameter values for the same classifier) that have similar thematic accuracy exhibit significantly different geometric properties (and, thus, global quality). For this reason, in many real classification problems, the quality of the maps obtained by the classification of VHR data is assessed also through a visual inspection. However, this procedure can provide just a subjective evaluation of the map quality that cannot be quantified. Thus, it is important to develop accuracy-assessment protocols for a precise, objective, and quantitative characterization of the quality of thematic maps in terms of both thematic and geometric properties [6]. These protocols could be used not only for assessing the quality of thematic maps generated by different classification systems but also for better driving the model selection of a single classifier, i.e., the selection of the optimum values for the free parameter of a supervised categorization algorithm.

An important area in which some studies related to the aforementioned problem have been done in the past is that of landscape ecology. Some approaches have been proposed in the landscape ecology literature to compare different maps by considering the spatial structure of the landscape [7] (and, thus, not only the thematic accuracy). As an example, in [8], different comparison methods that simultaneously consider both the spatial structure and the pixel-based overlap (i.e., the thematic accuracy) are presented. However, these methods are developed in a different framework and do not consider the particular properties of classification maps derived from VHR remote-sensing

images and the issues related to the tuning of the free parameters of a classifier.

In this paper, we address the aforementioned problem by proposing a novel protocol for a precise, automatic, and objective characterization of the accuracy of thematic maps derived from VHR images. The proposed protocol is based on the evaluation of two families of indices: 1) thematic accuracy indices and 2) a set of novel geometric indices that assess different properties of the objects recognized in the thematic map. The proposed protocol can be used in three different frameworks: 1) to objectively characterize the thematic and geometric properties of classification maps; 2) to select the map that better fits specific user-defined requirements; and 3) to identify the map that exhibits on average the best global properties if no specific requirements are defined. Moreover, we propose a novel approach for tuning the free parameters of supervised classification algorithms (e.g., SVM), which is based on the optimization of a multiobjective problem. The aim of this approach is to select the parameter values that result in a classification map that exhibits high geometric and thematic accuracies.

This paper is organized into six sections. The next section presents the background on the assessment of the thematic accuracy of land-cover maps. Section III describes the proposed accuracy-assessment protocol and discusses the two families of presented geometric and thematic indices. Section IV illustrates the proposed multiobjective criterion for the tuning of the free parameters (model selection) of a classifier. Section V presents the obtained experimental results, while Section VI draws the conclusion of this paper.

II. BACKGROUND ON THEMATIC ACCURACY ASSESSMENT OF CLASSIFICATION MAPS

In this section, we briefly recall the main concepts on the procedures used to assess the thematic accuracy of a classification map obtained by a supervised classifier [9], [10]. In general, the following two main issues should be addressed: 1) the collection of the labeled samples for both training and testing a supervised algorithm (which may require the subdivision of the reference-sample set into two or more disjoint sets) and 2) the choice of the statistical measure to evaluate the error (or accuracy) in pattern classification.

With respect to the first issue, several resampling methods have been proposed in the pattern-recognition and statistical literature, e.g., resubstitution, holdout, leave-one-out, cross-validation, and bootstrap [11]–[14]. Holdout is one of the most widely adopted resampling strategies in remote-sensing applications. It consists in partitioning the available labeled samples in two independent sets or in directly collecting two independent sets of samples in separate areas of the scene. One set is used for training the classifier, while the other one is used for assessing the classification accuracy. In some cases, it is preferable to split the available samples into three sets: 1) one for training the algorithm (*training set*); 2) one for tuning the free parameters of the classifier (*validation set*); and 3) one for assessing the final accuracy (*test set*). Holdout is less computationally demanding with respect to other methods (e.g., leave-one-out and k-fold cross-validation), and it is particularly

reliable when the available labeled samples are acquired in two spatially disjoint portions of the scene. Indeed, in this case, it is possible to assess the generalization capability of the classifier for test pixels that are spatially disjoint from the ones used for the training (which may present a different spectral behavior). With all the aforementioned resampling methods, it is important to adopt a stratified approach, i.e., the training and test sets (or each of the k folds) should contain approximately the same proportions of the class labels as the original data set. Otherwise, imbalanced and skewed results can be obtained.

With respect to statistical measures for accuracy evaluation, the complete description of the information that comes out from the comparison of the classification of test samples with the reference labeled data is given by the confusion (or error) matrix N. N is a square matrix of size $C \times C$ (where C is the number of information classes in the considered problem) defined as

$$N = \begin{bmatrix} n_{11} & n_{12} & n_{13} & \cdots & n_{1C} \\ n_{21} & n_{22} & \cdots & \cdots & \cdots \\ n_{31} & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ n_{C1} & \cdots & \cdots & \cdots & n_{CC} \end{bmatrix}. \quad (1)$$

The generic element n_{ij} of the matrix denotes the number of samples classified into category i ($i = 1, \ldots, C$) by the supervised classifier that are associated with label j ($j = 1, \ldots, C$) in the reference data set. This representation is complete, as the individual accuracy of each category is described along with both the errors of inclusion (commission errors) and errors of exclusion (omission errors) [9]. From the confusion matrix, different indices can be derived to summarize the information with a scalar value. Let us consider the sum of the elements of row i, $n_{i+} = \sum_{j=1}^{C} n_{ij}$ (which is the number of samples classified into category i in the classification map), and the sum of the elements of column j, $n_{+j} = \sum_{j=1}^{C} n_{ij}$ (which is the number of samples belonging to category j in the reference data set). Two commonly adopted indices are the overall accuracy (OA) and the kappa coefficient of accuracy ($kappa$), which are defined as

$$OA = \frac{\sum_{i=1}^{C} n_{ii}}{n} \quad (2)$$

$$kappa = \frac{n \sum_{i=1}^{C} n_{ij} - \sum_{i=1}^{C} n_{i+} n_{+i}}{n^2 - \sum_{i=1}^{C} n_{i+} n_{+i}} \quad (3)$$

where n is the total number of test samples. OA represents the ratio between the number of samples that are correctly recognized by the classification algorithm and the total number of test samples. The kappa coefficient of accuracy is a measure based on the difference between the actual agreement in the confusion matrix (as indicated by the main diagonal) and the chance agreement, which is indicated by the row and column totals (i.e., the marginals). The kappa coefficient is widely adopted,

as it uses also off-diagonal elements of the error matrix and compensates for chance agreement. However, as pointed out in [15], kappa statistics has also unfavorable features. The main objection to the kappa coefficient is that it was introduced as a measure of agreement for two observers (see [16]). Thus, the kappa coefficient evaluates the departure from the assumption that two observers' ratings are statistically independent rather than a measure of classification accuracy. For this reason, in [15], it is suggested to use other measures instead of kappa statistics, e.g., the class-averaged accuracy defined as

$$CA = \frac{1}{C} \frac{\sum_{j=1}^{C} n_{jj}}{n_{+j}} \qquad (4)$$

or an alternative coefficient based on the Kullback–Leibler information. We refer the reader to [9]–[11] for further details on accuracy-assessment procedures in remote-sensing image classification.

It is important to point out that all the aforementioned thematic accuracy measures do not consider the geometrical quality of the map under assessment and the shape of the objects that are present in the scene, thus resulting in the impossibility to assess the correctness of the geometry of the objects recognized by the classification algorithm. This is reasonable to evaluate the quality of the classification maps obtained by medium- or low-resolution images, where the geometry of the objects is difficult to characterize. On the contrary, for adequately assessing the quality of the classification maps obtained by VHR images, it is important to define indices that are capable to evaluate the geometrical properties of the maps and to use them together with more traditional thematic indices.

III. PROPOSED PROTOCOL FOR ACCURACY ASSESSMENT IN VHR IMAGES

In this section, we present the proposed protocol for accuracy assessment that is based on the computation of both thematic and geometric indices. The proposed procedure for thematic accuracy assessment is a simple refinement of the more traditional procedures described in the previous section, which takes into account particular properties of the classification of VHR images. On the contrary, the introduction of geometric indices to characterize the properties of the objects that are present in VHR images is one of the main contributions of this paper. Thematic and geometric indices are described in the following two subsections, respectively.

A. Thematic Error Indices

When VHR images are considered, we can clearly identify two different contributions to the overall thematic accuracy: 1) the accuracy obtained on homogeneous areas, where pixels are characterized by the spectral signature of only one class, and 2) the accuracy obtained on borders of the objects and details, where pixels are associated with a mixture of spectral signatures of different classes. These two contributions model the attitude of a classifier to correctly classifying homogeneous regions and
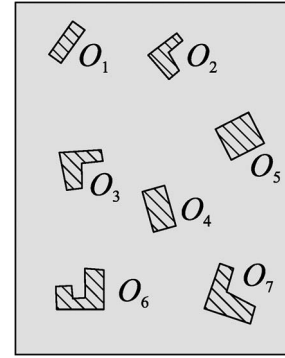

Fig. 1. Example of a map of reference objects.

high-frequency areas, allowing a more precise assessment of the quality of the classification map. The classification of mixed pixels is a difficult task with *crisp* classifiers, which should decide for the predominant class in the area associated with the pixel (*fuzzy* classifiers may be adopted in their place for considering the contributions of the different land-cover types to the spectral signature associated with each single pixel [17]). The proposed thematic accuracy assessment consists of the calculation of two separate indices: 1) thematic accuracy on homogeneous areas and 2) thematic accuracy on edge areas. This is accomplished by extending the holdout strategy by defining two independent test sets: one on homogeneous areas (pixel "inside" objects), and the other one on edge areas (pixels on the boundaries of objects). This results in the calculation of two independent confusion matrices. Any index derived from the confusion matrices (e.g., overall accuracy, kappa coefficient, etc.) may be adopted to calculate the accuracy on the two separate test sets. It is worth noting that different indices provide different information and can be used together (see the next section for a detailed discussion on the combined use of multiple indices for the tuning of the free parameters of a supervised classifier).

B. Geometric Error Indices

The geometric accuracy of a classification map is related to its precision in reproducing the correct geometry, the shapes, and the boundaries of the objects (e.g., buildings, streets, fields, etc.) that are present in the scene under investigation. In this paper, in order to quantify the geometric accuracy of maps characterized by very high spatial resolution, we define a set of object-based indices (error measures) that evaluate different geometric properties of the objects represented in a thematic map with respect to a reference map. Some of these indices are partially inspired to the measures used in the accuracy assessment of segmentation maps, while others are imported from different domains of image processing. These indices are computed by using a reference map that defines the exact shape, structure, and position of a set $\mathbf{O} = \{O_1, O_2, \ldots, O_d\}$ of d objects (e.g., buildings) adequately distributed in the scene under investigation and with different properties (see the example in Fig. 1). Generally, given the high resolution of VHR images, the map of reference objects can easily be defined by photointerpretation (few objects are sufficient for
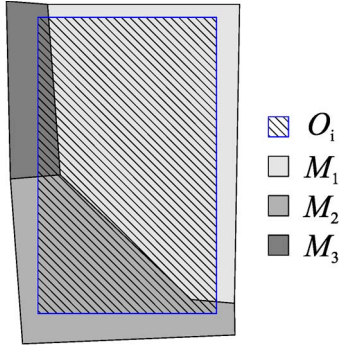
Fig. 2. Example of reference object O_i and the regions in the map that overlap with it. Region M_1 has the highest overlapping area with O_i and is selected according to (5).
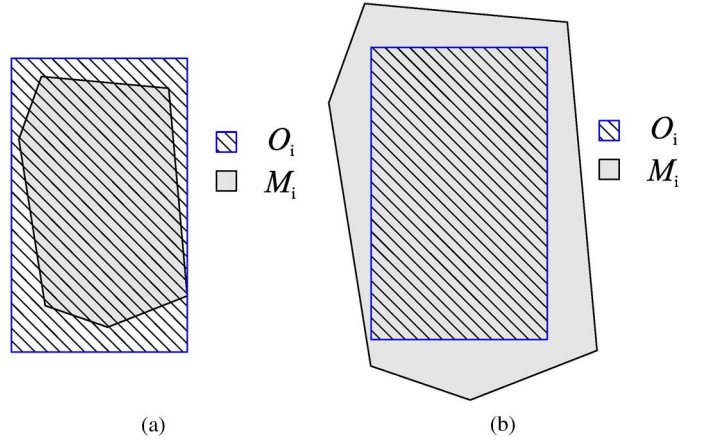

(a)   (b)

Fig. 3. (a) Example of oversegmentation. Region M_i that is recognized in the map is smaller than reference object O_i. (b) Example of undersegmentation. Region M_i that is recognized in the map is bigger than reference object O_i.

a good characterization of the properties of the map). Please note that the labels of the classes of the reference objects are not required for the computation of geometric accuracy indices. In this way, the evaluation of the geometric properties of the objects recognized in the map can be separated from the assessment of thematic accuracy. Moreover, we do not require having reference objects for all the classes considered in the classification problem, but only for the classes for which the geometric properties are important and the precise shape can easily be defined (e.g., buildings, fields, lakes, bridges, etc.).

Let us consider that the thematic map under assessment (e.g., obtained by an automatic algorithm or by photointerpretation) is made up of a set $\mathbf{M} = \{M_1, M_2, \ldots, M_r\}$ of r different regions of connected pixels (with 4- or 8-connectivity), such that each pixel in M_j, $j = 1, 2, \ldots, r$, is associated with the same label L_j, where L_j is one of the C information classes in $\mathbf{\Omega} = \{\omega_1, \omega_2, \ldots, \omega_C\}$. In order to calculate the geometric error measures, it is necessary to identify for each object O_i in the reference map the corresponding region in thematic map M_i. This can be done by considering the degree of overlapping between the pixels in reference object O_i and in regions M_j, $j = 1, 2, \ldots, k$. The region M_i in the map with the highest overlapping area with the object O_i (i.e., with the highest number of common pixels) is selected according to

$$M_i = \underset{\forall M_i \in \mathbf{M}}{\arg\max} |O_i \cap M_j| \qquad (5)$$

where $|\cdot|$ is the cardinality of a set and is used here to extract the number of pixels (area) from a region (see the example in Fig. 2). Given a pair (O_i, M_i), it is possible to calculate a set of local geometric error measures $err_i^{(h)}$, $i = 1, 2, \ldots, d$, $h = 1, 2, \ldots, m$, that evaluate the degree of mismatching (in terms of m different specific geometric properties) between the reference object and the corresponding region on the map. Global error measures $err^{(h)}$, $h = 1, 2, \ldots, m$, can then be defined on the basis of the local measures.

The adopted measures are as follows: 1) oversegmentation error; 2) undersegmentation error; 3) edge-location error; 4) fragmentation error; and 5) shape error.

1) Oversegmentation: Similar to the segmentation process, this error refers to the subdivision of a single object into several

distinct regions in the classification map [see the example in Fig. 3(a)]. The proposed local-error measure can be written as

$$OS_i(O_i, M_i) = 1 - \frac{|O_i \cap M_i|}{|O_i|}. \qquad (6)$$

This measure evaluates the ratio between the overlapping area of the two regions (O_i, M_i) and the area of the reference object. The index OS_i is defined in order to scale the output values in the [0, 1) range. The higher the value of the error, the higher the level of oversegmentation of the object O_i in the considered classification map. The value of this error is 0 in the optimal case where the two regions are in full agreement, while it tends to 1 in the worst case of just one common pixel among the two regions.

2) Undersegmentation: Undersegmentation refers to the classification errors that result in group of pixels belonging to different objects fused into a single region. The proposed local-error measure is defined as

$$US_i(O_i, M_i) = 1 - \frac{|O_i \cap M_i|}{|M_i|}. \qquad (7)$$

Unlike the oversegmentation error, the undersegmentation error is computed by considering the ratio between the area of overlapping among M_i and O_i and the area of the region on map M_i. Also, the US_i error varies in the [0, 1) range. Value 0 of this index corresponds to perfect agreement between M_i and O_i, while values that are close to 1 reflect a high amount of undersegmentation (i.e., region M_i is much bigger than the area of overlapping between regions M_i and O_i).

3) Edge Location: This index measures the precision of the object edges recognized in the classification map with respect to those of the actual object [see the example in Fig. 4(a)]. Let $e(O_i)$ denote the operator that extracts the set of edge pixels from a generic region O_i. In this framework, we consider the possibility to introduce a tolerance in the recognition of the object borders. This can be implemented by adopting an operator $e(\cdot)$ that extracts the border line of the objects with
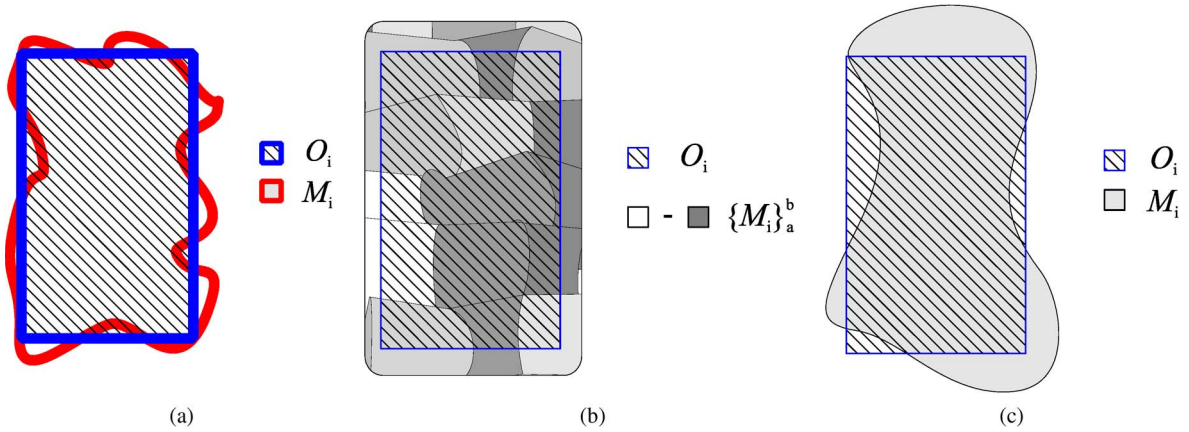
Fig. 4. Example of a region M_i that is recognized in the map (with corresponding reference object O_i) that exhibits (a) large edge-location error, (b) high level of fragmentation, and (c) relatively high shape error.

a width that is greater than 1 pixel (e.g., 2 or 3 pixels). The definition of the border error is given by

$$ED_i(O_i, M_i) = 1 - \frac{|e(O_i) \cap e(M_i)|}{|e(O_i)|}. \qquad (8)$$

This error measure varies in the [0, 1) range like the previous ones. A perfect matching in the borders of the two regions M_i and O_i leads to an error value that is equal to 0, whereas a large mismatching among the region edges results in error values that are close to 1.

4) Fragmentation Error: The fragmentation of a classification map refers to the problem of subpartitioning single objects into different small regions [see the example in Fig. 4(b)]. In order to quantitatively measure this type of error, we define a measure based on the number r_i of regions M_j, $j = 1, 2, \ldots, r_i$, that have at least one pixel in common with reference object O_i. For this reason, we define the set $\mathbf{R}_i$ of all the regions overlapping with reference object O_i as

$$\mathbf{R}_i = \{M_j, \forall j = 1, 2, \ldots, r_i : O_i \cap M_j \neq \varnothing\}. \qquad (9)$$

The proposed fragmentation error is then defined by the following:

$$FG_i(O_i, M_i) = \frac{r_i - 1}{|O_i| - 1}. \qquad (10)$$

This error value is scaled in the [0, 1] range. The value is 0 in the optimal case when only one region M_j is overlapping with reference object O_i, whereas it is 1 in the worst case where all the pixels of object O_i belong to different regions M_j on the map. The measure is normalized with respect to the size (area) of reference object O_i. It is worth noting that the fragmentation error is correlated with the oversegmentation error but differs from the latter because it takes into account all the r_i regions M_j that overlap with real object O_i, instead of the area of the single region M_i obtained by (5).

5) Shape Error: This error is used to evaluate the shape difference between an object O_i and the corresponding region M_j on the map [see the example in Fig. 4(c)]. In order to characterize the shape of an object, several shape factors have been proposed in the literature and can be adopted (e.g.,

compactness, sphericity, and eccentricity [18]). Thus, the shape error can be defined as the absolute value of the difference in the selected shape factor $sf(\cdot)$ of the two regions M_i and O_i

$$SH_i = \|sf(O_i) - sf(M_i)\|. \qquad (11)$$

It is worth noting that, by adopting shape factors normalized in the [0, 1] range, the defined shape-error measure will vary in the same range.

On the basis of the previously defined measures of local errors (i.e., errors associated with single objects in the map), it is then possible to estimate the global behaviors of the geometric properties of the classification map. Global error measurements can be obtained by averaging the local errors over the d measurements associated with the reference objects in $\mathbf{O}$, i.e., a generic global-error-measure characterizing property $err^{(h)}$ of the map can be expressed as

$$err^{(h)} = \frac{1}{d} \sum_{i=1}^{d} err_i^{(h)} \qquad (12)$$

where $err_i^{(h)}$ is a local error h on object i. In this way, we give the same weight to the errors over the d objects, independently from their size. Other possible definitions of the global measures may take into account the size of the different objects, i.e.,

$$err^{(h)} = \frac{1}{d} \sum_{i=1}^{d} |O_i| err_i^{(h)} \qquad (13)$$

or can weight differently the objects on the basis of specific user-defined requirements, i.e.,

$$err^{(h)} = \frac{1}{d} \sum_{i=1}^{d} \lambda_i \, err_i^{(h)} \qquad (14)$$

where λ_i, $i = 1, 2, \ldots, n$, are defined by the user. For example, the user may specify that geometric errors on buildings are more important than geometric errors on other objects, like streets, crop fields, or lakes. Global measures are then used to estimate different geometric properties of the map. Combining

the different global indices in a single measure that averages geometric indices is also possible. Nevertheless, this procedure would result in a measure that is difficult to understand.

IV. PROPOSED MULTIOBJECTIVE STRATEGY FOR CLASSIFIER-PARAMETER OPTIMIZATION

Other than the quality assessment of the classification maps obtained according to different procedures (e.g., different automatic classifiers, photointerpretation, etc.), an accuracy index is also an important measure for tuning the free parameters of supervised classifiers (this process is also indicated as model selection). Let us consider a generic supervised algorithm for which a vector θ of free parameters should be selected in order to optimize the quality of the output map. Standard approaches are based on the adoption of a scalar index to assess the thematic accuracy of the map (e.g., the overall accuracy or the kappa coefficient) and on the selection of vector θ that maximizes such a scalar value on the test samples. If a vector I of quality indices that characterize different thematic and geometric properties of the classification map is considered, the selection of θ should be based on a different optimization strategy. The simplest (yet empirical and only partially reliable) strategy is to define a single error function $E(\cdot)$ combining the m proposed error measures according to a weighted average

$$E(\mathbf{O}, \mathbf{M}) = \sum_{j=1}^{m} c_j \, err^{(j)} \tag{15}$$

where the terms c_j, $j = 1, 2, \ldots, m$, are defined by the user. The set of parameter values of θ that produces the classification map that minimizes $E(\mathbf{O}, \mathbf{M})$ represents the solution to the considered problem. Nevertheless, this formulation has an important drawback: The definition of c_j (which significantly affects the final result) is very critical because of the different intrinsic scales of the considered errors. In addition, the physical information conveyed by the resulting global index is difficult to understand.

To overcome this drawback, we propose to model our problem as a multiobjective minimization problem, where the multiobjective function $\mathbf{g}(\theta)$ is made up of m different objectives $g_1(\theta), g_2(\theta), \ldots, g_m(\theta)$ that represent the set of adopted error measures computed for different values of the classifier parameters (e.g., different thematic and geometric indices). All the different objectives of $\mathbf{g}(\theta)$ have to be jointly minimized and are considered equally important. In general, all the proposed thematic indices (evaluated on homogeneous and border areas with different statistical parameters) and geometric indices could be used for the definition of $\mathbf{g}(\theta)$. However, depending on the application, it could be more appropriate to use different subsets of the presented indices as objectives of the optimization problem (e.g., for meeting some particular quality properties of the classification map required by the end users). Thus, the multiobjective problem can be formulated as follows:

$$\min_{\theta \in S} \{\mathbf{g}(\theta)\}, \mathbf{g}(\theta) = [g_1(\theta), g_2(\theta), \ldots, g_m(\theta)]$$

$$\text{subject to } \theta = (\theta_1, \theta_1, \ldots, \theta_h) \in S \subseteq \mathbb{R}^h \tag{16}$$
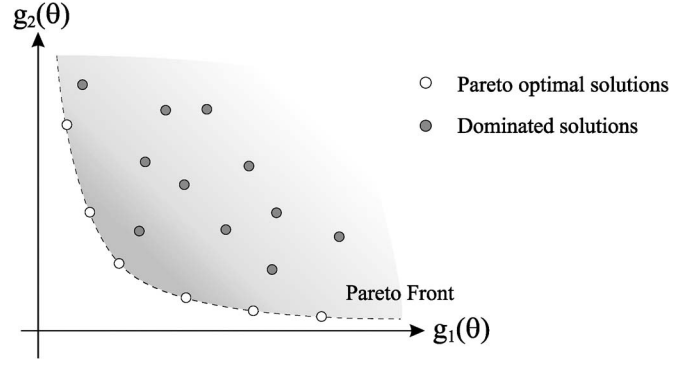

Fig. 5. Example of Pareto-optimal solutions and dominated solutions in a two-objective search space.
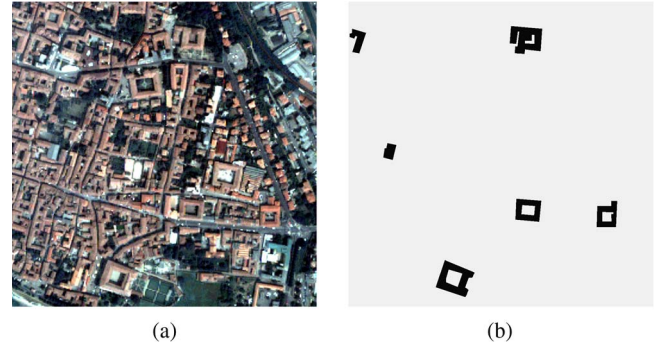

Fig. 6. (a) Real color composition of the image acquired by the QuickBird satellite in the city of Pavia (northern Italy). (b) Map of reference objects.

where S denotes the search space for the classifier parameters. This problem is characterized by a vector-valued objective function $\mathbf{g}(\theta)$ and cannot be solved in order to derive a single solution like in optimization problems characterized by a single objective function. Instead, a set of optimal solutions P^* can be obtained by following the concept of Pareto dominance. In greater detail, a solution θ^* is said to be Pareto optimal if it is not dominated by any other solution in the search space, i.e., there is no other θ such that $g_i(\theta) \leq g_i(\theta^*)$ ($\forall i = 1, 2, \ldots, m$) and $g_j(\theta) < g_j(\theta^*)$ for at least one j ($j = 1, 2, \ldots, m$). This means that θ^* is Pareto optimal if there exists no other subset of classifier parameters θ that would decrease an objective without simultaneously increasing another one (Fig. 5 clarifies this concept with a graphical example). The set P^* of all optimal solutions is called Pareto-optimal set. The plot of the objective function of all solutions in the Pareto set is called Pareto front $PF^* = \{\mathbf{g}(\theta) | \theta \in P^*\}$. The main advantage of the multiobjective approach is that it avoids to aggregate metrics capturing multiple objectives into a single measure. On the contrary, it allows one to effectively identify different possible tradeoffs between maps exhibiting different thematic and geometric properties.

Because of the complexity of the search space, an exhaustive search of the set P^* of optimal solutions is unfeasible. Thus, instead of identifying the true set of optimal solutions, we aim to estimate a set $\hat{P}^*$ of nondominated solutions with objective values as close as possible to the Pareto front. This estimation can be done with different multiobjective optimization algorithms (e.g., multiobjective evolutionary algorithms [19], [20]).

TABLE I
NUMBER OF SAMPLES IN THE TRAINING AND TEST SETS (PAVIA DATA SET)

Class	Number of patterns		
	Training set	Test set on edge areas	Test set on homogeneous areas
Water	180	55	150
Tree areas	348	95	250
Grass areas	323	90	160
Roads	984	182	381
Shadow	750	297	325
Red buildings	2271	442	1040
Gray buildings	602	167	250
White building	275	98	100
TOTAL	5733	1426	2656

The final selection of the optimal solution among all estimated nondominated solutions is demanded to the user, who can select the best tradeoff among the considered objectives on the basis of the specific application (e.g., one could tolerate to have undersegmented maps rather than oversegmented ones or prefer to have less fragmented objects rather than high precision in the shape, etc.).

V. EXPERIMENTAL RESULTS

This section presents an experimental analysis aimed at studying the reliability of the proposed protocol for the accuracy assessment of the classification maps obtained by VHR images. We first applied the proposed indices to the quality assessment of different thematic maps obtained by the classification (carried out with different automatic techniques) of a QuickBird image acquired in the city of Pavia, Italy. Then, in a second set of experiments, we applied the proposed multiobjective strategy to the model selection of an SVM classifier in the analysis of a different QuickBird image acquired in the city of Trento, Italy. In our implementation of geometric indices, we considered a tolerance of 3 pixels for the edge-location error, and we selected eccentricity [18] as shape factor for the evaluation of the shaper error. The global geometric errors were computed on the basis of (12).

A. Quality Assessment of Classification Maps

The first considered data set is made up of a QuickBird multispectral image acquired in the city of Pavia (northern Italy) on June 23, 2002. In particular, we used a panchromatic image and a pan-sharpened multispectral image [see Fig. 6(a)] obtained by applying a Gram–Schmidt fusion technique [21] to the panchromatic channel and to the four bands of the multispectral image. The image size is 1024×1024 pixels with a spatial resolution of 0.7 m. Greater details about this data set can be found in [1]. Table I presents the number of labeled reference samples for each set and class. The test pixels used for the assessment of thematic accuracy were collected on both edge and homogeneous areas. Test set pixels were taken from areas of the scene spatially disjoint from those related to the training samples. Fig. 6(b) shows the map of the reference objects used for the evaluation of geometric error indices. In

particular, six different buildings were manually selected and considered as reference objects. It is worth noting that, given the VHR of the images, the procedure for digitizing few reference objects is simple and very fast.

In our experiments, we obtained different thematic maps of the scene by using different automatic classification systems. The different systems were defined by varying the feature vector (i.e., considering only spectral features and/or also multiscale/multilevel contextual features), the supervised classification algorithms (i.e., parallelepiped, maximum-likelihood (ML), and SVM classifiers), and, in some cases, adding a postprocessing phase for regularizing the final classification map. These systems were chosen with the goal to obtain classification maps with different properties. Fig. 7 shows the thematic maps obtained by the different considered classification systems. In particular, the (a)–(d) maps are obtained by considering a feature vector that is made up of only the original spectral features. Map (a) is obtained by using a very simple parallelepiped classifier (with $\sigma = 2$) [22], map (b) is derived by adopting a Gaussian ML classifier, map (c) is obtained by applying a majority filter (with a sliding window of size 3×3) as postprocessing to map (b) [22], and map (d) is the result of classification with SVM (using Gaussian kernels). The (e)–(h) maps are yielded using both spectral and contextual features and adopting SVM as the classification algorithm. Map (e) is obtained considering the features extracted on the basis of the generalized Gaussian pyramid decomposition. In detail, the images were iteratively analyzed by a Gaussian kernel low-pass filter (with 5×5 square analysis window) and were undersampled by a factor of two. We exploited five levels of pyramidal decomposition to characterize the spatial context of pixels and to label each pixel of the scene under investigation. Maps (f)–(h) are obtained using the multilevel context-based feature-extraction approach proposed in [1]; different statistical parameters are extracted from the pixels in each region defined at six different levels by a hierarchical-segmentation process. In particular, for map (f), we considered the mean value for the first five levels and the standard deviation for levels 3, 4, and 5, while for map (g), we considered only the mean for all first five levels. Map (h) is obtained considering the mean value extracted from all six segmentation levels.

Tables II and III report the thematic accuracies and the geometric error indices associated with the obtained maps,
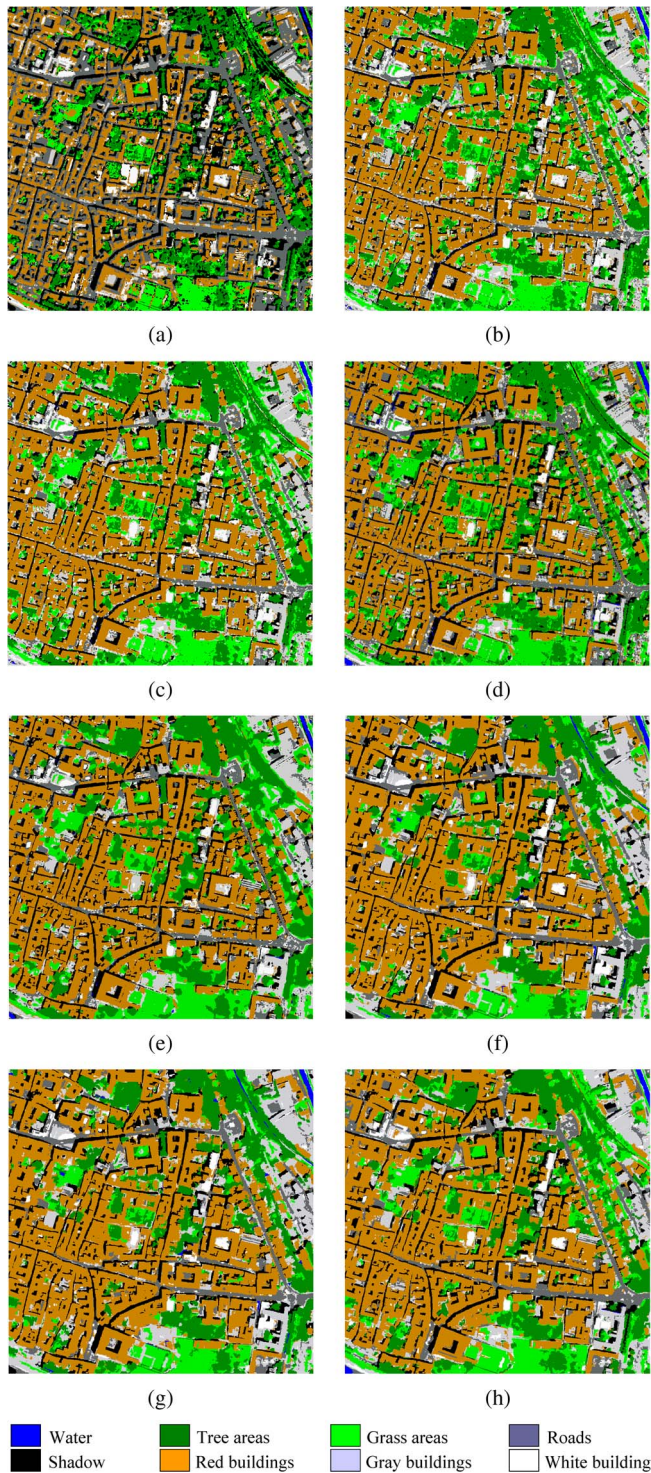
Fig. 7. Thematic maps obtained by different classification systems applied to the Pavia QuickBird image. (a) Parallelepiped. (b) ML. (c) ML with postprocessing. (d) SVM. (e) SVM Gaussian pyramid. (f) SVM multilevel features—five levels (1). (g) SVM multilevel features—five levels (2). (h) SVM multilevel features—six levels.

respectively. Considering the eight different maps, we can easily observe that, as expected, the thematic maps obtained by pixel-based classification approaches [maps (a), (b), and (d)] are less accurate than those obtained by context-based approaches. This general behavior is clearly pointed out also by thematic accuracy indices. The geometric error measurements give us important additional information about the different properties of the maps. In particular, we note that the maps obtained by pixel-based approaches are generally more over-segmented and fragmented than the maps obtained by context-based classification systems, but they have also the important property to be less undersegmented. In the considered scene, we can observe that the buildings are very close to each other. Thus, most of the considered classifiers merge regions associated with distinct objects (i.e., buildings) into a single region. The aforementioned problem is captured by the proposed geometric indices, which indicates that most of the maps have an undersegmentation error that is higher than the oversegmentation error [except for map (a)]. This problem strongly affects also the recognition of the correct shape of the objects. For this reason, we can observe that, on this data set, the shape error is highly correlated with the undersegmentation error. We can further observe that the edge-location error is, in general, quite high for all the obtained maps (even if a tolerance of 3 pixels is considered). This indicates that the considered classification techniques can scarcely model the correct borders of the objects.

Analyzing the single maps, we can observe that map (a) has very low quality in terms of thematic accuracy and in terms of most of the geometric indices. In particular, this map is sharply oversegmented and fragmented, as indicated by the geometric errors; this is confirmed by a visual inspection. Map (b) has better quality than map (a): it exhibits higher thematic accuracy (both on homogeneous and border areas) and better geometric properties in terms of undersegmentation and border error. Map (c) [obtained by a postprocessing applied to map (b)] results in slightly higher thematic accuracy and in smaller oversegmentation, fragmentation, and border errors than map (b). Nevertheless, the majority postprocessing leads to a slight increase in the undersegmentation error. Map (d) is the most accurate among those obtained with a pixel-based approach: This is pointed out by both thematic and geometric indices. In particular, this map exhibits the smallest undersegmentation and edge-location errors among all considered maps. Map (e) exhibits important advantages with respect to the aforementioned maps, showing smaller oversegmentation and fragmentation errors as well as higher thematic accuracies. Nevertheless, the thematic accuracies (particularly on border areas) are smaller than those of maps (f)–(h). The geometric indices result to be particularly important for the characterization of the different maps obtained by the multilevel feature-extraction technique [maps (f)–(h)], which have high and very similar thematic accuracies. Map (f) is the most accurate from a thematic point of view, but maps (g) and (h) exhibit better geometric characteristics (e.g., undersegmentation and edge-location errors) than map (f). As it is possible to observe in Fig. 8, map (f) is affected by undersegmentation problems, as it merges different objects in the same region. On the contrary, map (h) correctly models the different buildings. This difference is clearly pointed out by the values of the undersegmentation error. Thus, considering both thematic and geometric indices, we can select map (h) as more reliable than map (f) (which would be preferred considering only thematic accuracies) because it presents a better tradeoff among different properties of the maps. It is worth noting

TABLE II
THEMATIC ACCURACIES COMPUTED ON THE TEST SET ON HOMOGENEOUS AND EDGE AREAS AND ON BOTH OF THEM (COMPLETE TEST SET)
EVALUATED IN TERMS OF OVERALL ACCURACY (OA) AND KAPPA COEFFICIENT ($kappa$) (PAVIA DATA SET)

Map	Complete Test set (homog. + edge areas)		Test set on homogeneous Areas		Test set on edge Areas	
	OA%	*kappa*	*OA%*	*kappa*	*OA%*	*kappa*
(a) Parallelepiped	63.6%	0.564	77.7%	0.725	37.4%	0.287
(b) ML	83.0%	0.789	94.1%	0.925	62.4%	0.543
(c) ML post-processing	84.4%	0.805	95.2%	0.939	64.2%	0.564
(d) SVM	84.2%	0.801	94.7%	0.932	64.6%	0.563
(e) SVM Gaussian Pyramid	86.3%	0.828	95.0%	0.936	70.1%	0.631
(f) SVM Multilevel 5 levels (1)	90.0%	0.874	96.8%	0.960	77.1%	0.716
(g) SVM Multilevel 5 levels (2)	88.9%	0.861	97.1%	0.963	73.6%	0.677
(h) SVM Multilevel 6 levels	89.3%	0.866	96.2%	0.952	76.5%	0.711

TABLE III
GEOMETRIC ERROR INDICES (PAVIA DATA SET)

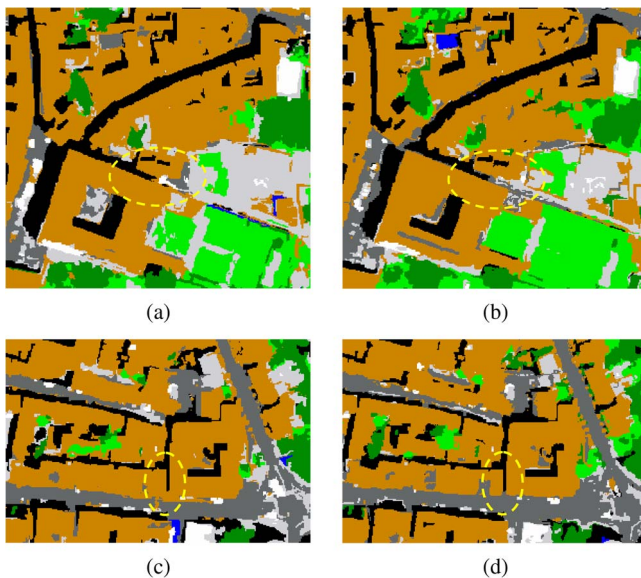| Map	Under-segmentation	Over-segmentation	Edge location	Fragmentation	Shape
(a) Parallelepiped	26.9 %	44.6 %	77.4 %	27.6 %	13.8 %
(b) ML	26.2 %	9.7 %	66.9 %	9.4 %	14.5 %
(c) ML post-processing	30.2 %	8.5 %	68.0 %	8.2 %	16.2 %
(d) SVM	16.9 %	12.7 %	58.4 %	7.4 %	12.9 %
(e) SVM Gaussian Pyramid	29.3 %	6.3 %	64.2 %	4.7 %	16.8 %
(f) SVM Multilevel 5 levels (1)	47.6 %	4.1 %	74.1 %	3.1 %	24.8 %
(g) SVM Multilevel 5 levels (2)	26.8 %	6.2 %	62.4 %	5.9 %	19.2 %
(h) SVM Multilevel 6 levels	27.1 %	4.8 %	58.5 %	4.3 %	17.0 %

(a)   (b)

(c)   (d)

Fig. 8. Details of the thematic maps. (a)–(c) Undersegmentation problems in map (f). (b)–(d) Correct recognition of distinct buildings in map (h) (Pavia data set). (a) Detail of map (f). (b) Detail of map (h). (c) Detail of map (f). (d) Detail of map (h).
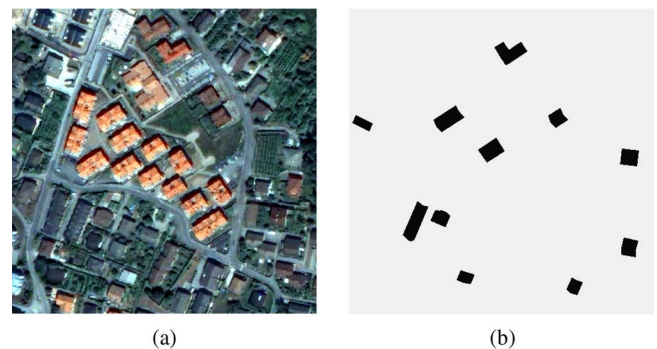

(a)   (b)

Fig. 9. (a) Real color composition of the multispectral image acquired by the QuickBird satellite in the city of Trento (northern Italy). (b) Map of reference objects.

In general, the selection of the highest quality map depends on the kind of application and/or on end-user requirements. In this context, the proposed indices are a valuable tool that can drive the selection of the best thematic map in accordance to the application constraints.

B. Multiobjective Strategy for the Model Selection of Supervised Algorithms

In the second set of experiments, we used the proposed multiobjective technique for the model selection of an SVM

that the property of correctly recognizing and distinguishing single objects in the scene can be very important for urban-area analysis, specifically in applications like building detection.

TABLE IV
NUMBER OF SAMPLES IN THE TRAINING AND TEST SETS (TRENTO DATA SET)

Class	Number of patterns		
	Training set	Test set on edge areas	Test set on homogeneous areas
Roads	58	63	47
Red roof buildings	71	70	73
Dark roof buildings	68	51	66
Bright roof buildings	39	38	39
Shadow	43	46	40
Vegetation	88	57	83
TOTAL	367	325	348

TABLE V
THEMATIC AND GEOMETRIC ACCURACIES/ERROR INDICES OF SEVEN SOLUTIONS SELECTED AMONG ALL PARETO-OPTIMAL POINTS
ESTIMATED BY THE GENETIC ALGORITHM. EACH SELECTED SOLUTION EXHIBITS AN ACCURACY INDEX THAT HAS THE HIGHEST
VALUE AMONG ALL SOLUTIONS (EXPERIMENTS WITH SEVEN ERROR INDICES IN THE OPTIMIZATION PROBLEM)

Map	SVM parameters		Thematic accuracies		Geometric errors				
	C	$2\sigma^2$	kappa homog.	kappa edge	Under-segmentation	Over-segmentation	Edge location	Fragm.	Shape
(2a)	3187	0.820	0.951	0.892	12.7%	21.9%	51.3%	13.3%	14.7%
(2b)	4032	6.094	0.926	0.930	11.0%	25.5%	59.4%	15.4%	14.0%
(2c)	92	6.725	0.909	0.911	7.3%	29.9%	56.6%	13.1%	11.2%
(2d)	3464	0.221	0.926	0.771	31.9%	19.6%	63.6%	12.4%	16.8%
(2e)	2119	4.634	0.930	0.922	10.9%	25.2%	50.0%	15. 5%	13.9%
(2f)	4725	0.617	0.930	0.782	23.3%	21.4%	59.4%	10.5%	18.4%
(2g)	86	6.767	0.905	0.903	9.4%	30.2%	52.4%	12.7%	11.0%

classifier with radial basis function Gaussian kernels [23], [24]. The free parameters of the classifier are the regularization term C and the spread σ^2 of the Gaussian kernel. The experiments were carried out on a VHR image acquired by the QuickBird multispectral scanner on an urban area in the south of the city of Trento (Italy) in July 2006 [see Fig. 9(a)]. We used a panchromatic image and a pan-sharpened multispectral image obtained by applying a Gram–Schmidt fusion technique to the panchromatic channel and to the four bands of the multispectral image. The image size is 500×500 pixels with a spatial resolution of 0.7 m. From the panchromatic and pan-sharpened multispectral bands, we extracted textural features by applying an occurrence filter with 5×5 window size and computing mean, data range, and variance. Thus, the final feature vector is made up of 20 features (5 spectral features and 15 textural features). The available set of reference samples included a training set, a test set on homogeneous areas, and a test set on border areas. The following six classes were considered: 1) roads; 2) red buildings; 3) dark buildings; 4) bright buildings; 5) shadow; and 6) vegetation.

Table IV presents the number of labeled reference samples for each set and class. Fig. 9(b) shows the map of the 11 reference objects used for the evaluation of geometric error indices.

The strategy for the model selection proposed in Section IV can be applied considering different sets of thematic and geometric indices as objectives of the optimization problem, depending on the specific application. In our analysis, we performed two sets of experiments considering the following: 1) seven objectives (two thematic and five geometric error
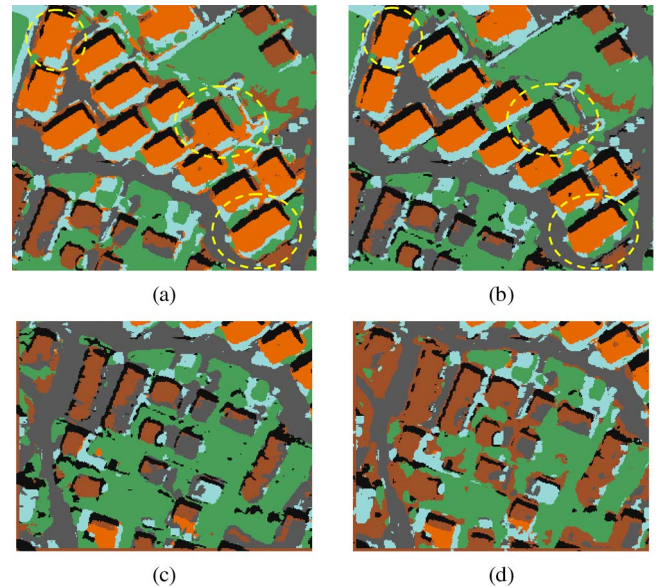

Fig. 10. Details of the maps associated with different Pareto-optimal solutions (experiments with seven error indices in the optimization problem; Trento data set). (a) Detail of map (2a). (b) Detail of map (2g). (c) Detail of map (2c). (d) Detail of map (2d).

indices) and 2) two objectives (one thematic and one geometric error indices). These two sets of experiments represent examples of the practical use of the proposed multiobjective approach in real problems, but any other combination of thematic and geometric indices may be used in the optimization problem for parameter tuning.

TABLE VI
PARETO-OPTIMAL SOLUTIONS ESTIMATED BY THE GENETIC ALGORITHM FOR THE EXPERIMENT WITH TWO ERROR INDICES (TRENTO DATA SET)

Map	SVM parameters		Error indices	
	C	$2\sigma^2$	1-kappa (homog. areas)	Under-segmentation error
(3a)	440	4.387	6.31%	9.95%
(3b)	155	5.836	6.66%	9.43%
(3c)	449	4.672	5.96%	10.36%
(3d)	799	1.177	5.61%	13.23%
(3e)	12	6.212	8.77%	6.86%
(3f)	24	6.455	7.01%	7.89%
(3g)	665	1.146	4.91%	13.62%
(3h)	667	1.145	5.26%	13.49%
(3i)	12	6.218	8.42%	7.19%
(3l)	19	6.059	7.71%	7.77%

1) Experiments with Seven Error Indices in the Optimization Problem: In this set of experiments, we defined model selection as a multiobjective optimization problem made up of seven objectives: the five geometric measures presented in Section III (i.e., undersegmentation, oversegmentation, edge-location, fragmentation, and shape errors) and the two thematic errors based on kappa coefficient (calculated as $1 - kappa$) on the homogeneous and border test sets. Please note that, in our experimental analysis, we used a thematic error index based on the popular kappa coefficient, but any other index may be used in its place (e.g., the overall error). For the estimation of Pareto-optimal solutions, we adopted a genetic multiobjective algorithm (a variation of NSGA-II) [25]. The population size and the maximum number of generations were set to 30 and 20, respectively. Among all Pareto-optimal solutions obtained by the genetic algorithm, we selected seven solutions (used as an example in this discussion), characterized by different tradeoffs among the different indices (see Table V). The selected solutions are characterized by the lowest error among all solutions for each index [e.g., map (2a) presents the highest thematic accuracy on homogeneous areas, map (2b) exhibits the highest thematic accuracy on edge areas, map (2c) shows the minimum undersegmentation error, etc.].

All these Pareto-optimal solutions are associated with maps having different thematic and geometric properties. For example, Fig. 10 shows some details of maps (2a)–(2g) and (2c)–(2d). Map (2a) [Fig. 10(a)] exhibits the highest kappa coefficient of accuracy on homogeneous areas, but the shape of red-roof buildings is not well recognized. On the contrary, map (2g) [Fig. 10(b)] has a smaller thematic accuracy but better models the shape of the buildings. This behavior can also be observed by a visual inspection of the maps. Map (2c) [Fig. 10(c)] has the lowest undersegmentation error, whereas map (2d) [Fig. 10(d)] has good oversegmentation properties, in spite of significant undersegmentation errors (which also affect the recognition of the shape of the objects).

2) Experiments with Two Error Indices in the Optimization Problem: In this second set of experiments, two objectives were considered in the optimization problem: 1) the kappa coefficient of accuracy on homogeneous areas and 2) the undersegmentation error. This represents an example in which we would like to select the SVM model that results in classification maps with the best tradeoff among thematic accuracy and precision
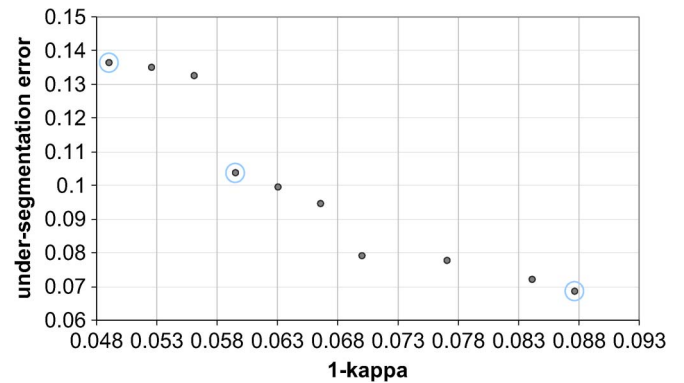

Fig. 11. Estimated Pareto-optimal solutions for the experiment with two error indices in the optimization problem.

in detecting separate buildings (undersegmentation error). The genetic algorithm adopted for the estimation of the Pareto front resulted in the estimation of the ten optimal solutions reported in Table VI.

Fig. 11 shows the estimated Pareto front. The selection of one model for the SVM classifier (i.e., the values of C and $2\sigma^2$) depends on the requirements of the specific application. For example, we selected three possible models from the Pareto-optimal solutions that leads to: i) the map with the highest kappa coefficient of accuracy on homogeneous areas [map (3g)]; ii) the map with the lowest under-segmentation error [map (3e)]; iii) a good tradeoff between the two competing objectives [map (3c)]. A qualitative visual analysis of the obtained maps confirms that map (3g) [Fig. 12(a)] has some under-segmentation problems (but it has the smallest possible under-segmentation error for the obtained kappa value), map (3e) [Fig. 12(b)] is less under-segmented (and exhibits the highest possible kappa accuracy for the value of the obtained under-segmentation error), and map (3c) [Fig. 12(c)] can be considered a good tradeoff between the two considered objectives.

It is worth noting that different error indices can be included in the multiobjective model selection. The choice of error indices should reflect the properties that the end users desire to optimize in the classification map. Other experiments, carried out using different error indices, confirmed the reliability of the proposed multiobjective model-selection technique based on the proposed accuracy-assessment protocol.
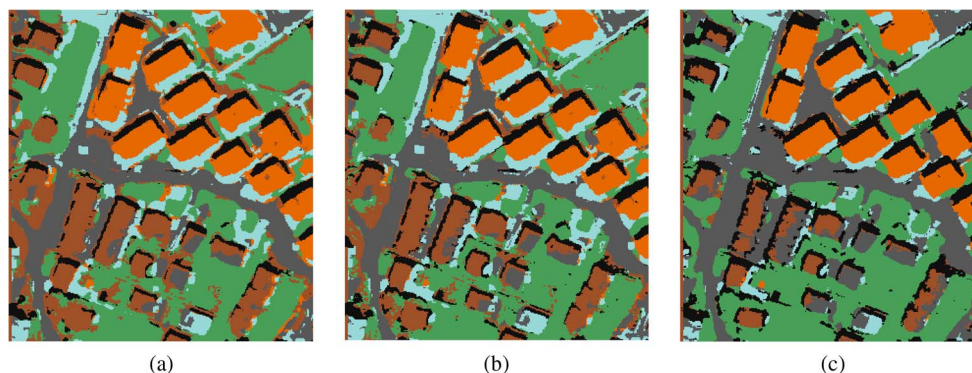
Fig. 12. Details of the maps associated with the three selected solutions (experiment with two error indices in the optimization problem; Trento data set). (a) Detail of map (3g). (b) Detail of map (3c). (c) Detail of map (3e).

VI. DISCUSSION AND CONCLUSION

In this paper, a novel protocol for the accuracy assessment of the thematic maps obtained by the classification of VHR images has been presented. The proposed protocol is based on the evaluation of a set of error measures that can model the thematic and geometric properties of the obtained map. In particular, we presented a set of indices that characterize five different types of geometric errors in the classification map: 1) oversegmentation; 2) undersegmentation; 3) edge location; 4) shape distortion; and 5) fragmentation. The proposed geometric measures can be jointly used with the traditional thematic accuracy measures for a precise characterization of the properties of a thematic map derived by VHR images. The presented protocol can be used in three different frameworks: 1) assessing the quality of a classification map in an automatic, objective, and quantitative way; 2) selecting the classification map, among a set of different maps, that is more appropriate for the specific application on the basis of user-defined requirements; and 3) selecting the values of the free parameters of a supervised classification algorithm that result in the most appropriate classification map. Regarding this latter point, we have introduced a new technique for tuning the free parameters of supervised classifiers that is based on the optimization of a multiobjective problem, which results in parameter values that jointly optimize thematic and geometric error indices on the classification map.

The experimental results obtained on two VHR images confirm that the proposed geometric indices can accurately characterize the properties of classification maps, providing objective and quantitative error measures, which are in agreement with the observations derived by a visual inspection of the considered maps. Moreover, the proposed approach for tuning the free parameters of supervised classifiers resulted effective in the selection of the free parameters of SVM classifiers. This approach allows one to better characterize the tradeoff among the different thematic and geometric indices and to select the model in accordance with user requirements and application constraints.

It is worth noting that the proposed approach represents a step toward a new direction in accuracy assessment of the classification maps derived from VHR images. However, some issues need to be further studied. One open issue is related to the definition of reference objects and the evaluation of geometric indices in the case of adjacent objects that cannot be easily separated (e.g., different overlapped tree crown). Other issues are related to the definition of additional geometric indices to include in the proposed protocol and multiobjective strategy for considering different geometric properties. Moreover, as additional future developments of this paper, we plan to extend the proposed multiobjective approach based on the evaluation of both thematic and geometric indices to the tuning of other variables of the classification system (not only related to the classification algorithm), e.g., for selecting the features to be given as input to the classifier or the parameters defining a post-processing, which strongly impact on the geometric properties of the final classification map.

REFERENCES

[1] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2587–2600, Sep. 2006.

[2] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.

[3] J. Chanussot, J. A. Benediktsson, and M. Fauvel, "Classification of remote sensing images from urban areas using a fuzzy possibilistic model," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 40–44, Jan. 2006.

[4] P. Gamba, F. Dell'Acqua, G. Lisini, and G. Trinni, "Improved VHR urban area mapping exploiting object boundaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 8, pp. 2676–2682, Aug. 2007.

[5] R. Bellens, S. Gautama, and L. Martinez-Fonte, "Improved classification of VHR images of urban areas using directional morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2803–2813, Oct. 2008.

[6] L. Bruzzone and C. Persello, "A novel protocol for accuracy assessment in classification of very high resolution multispectral and SAR images," in *Proc. IEEE IGARSS*, Boston, MA, July 6–11, 2008, vol. 2, pp. II-265–II-268.

[7] R. White, "Pattern based comparisons," *J. Geogr. Syst.*, vol. 8, no. 2, pp. 145–164, May 2006.

[8] A. Hagen-Zanker, "Map comparison methods that simultaneously address overlap and structure," *J. Geogr. Syst.*, vol. 8, no. 2, pp. 165–185, May 2006.

[9] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data*. Boca Raton, FL: Lewis Publ., 1999.

[10] G. M. Foody, "Status of land cover classification accuracy assessment," *Remote Sens. Environ.*, vol. 80, no. 1, pp. 185–201, Apr. 2002.

[11] A. Baraldi, L. Bruzzone, and P. Blonda, "Quality assessment of classification and cluster maps without ground truth knowledge," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 857–873, Apr. 2005.

[12] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. IJCAI*, 1995, pp. 1137–1145.

[13] B. Efron, "Estimating the error rate of a prediction rule: Improvement on cross-validation," *J. Amer. Stat. Assoc.*, vol. 78, no. 382, pp. 316–331, Jun. 1983.

[14] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical patter recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.

[15] R. Nishii and S. Tanaka, "Accuracy and inaccuracy assessments in land-cover classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 491–498, Jan. 1999.

[16] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Phychol. Meas.*, vol. 20, no. 1, pp. 37–40, Apr. 1960.

[17] F. Wang, "Fuzzy supervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 2, pp. 194–201, Mar. 1990.

[18] K. R. Castleman, *Digital Image Processing*. Upper Saddle River, NJ: Prentice-Hall, 1996, pp. 492–498.

[19] C. M. Fonseca and P. J. Fleming, "Multiobjective optimization and multiple constraint handling with evolutionary algorithms—Part I: A unified formulation," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 28, no. 1, pp. 26–37, Jan. 1998.

[20] A. Konak, D. W. Coit, and A. E. Smith, "Multi-objective optimization using genetic algorithms: A tutorial," *Reliab. Eng. Syst. Safety*, vol. 91, no. 9, pp. 992–1007, Sep. 2006.

[21] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6 011 875, Jan. 4, 2000.

[22] *ENVI User Guide*. [Online]. Available: http://www.ittvis.com/portals/0/pdfs/envi/Reference_Guide.pdf

[23] M. Dalponte, M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1416–1427, May 2008.

[24] L. Bruzzone, C. Mingmin, and M. Marconcini, "A novel transductive SVM for semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.

[25] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. Chichester, U.K.: Wiley, 2001.

Claudio Persello (S'07) received the Laurea (B.S.) and Laurea Specialistica (M.S.) degrees in telecommunication engineering from the University of Trento, Trento, Italy, in 2003 and 2005, respectively, where he is currently working toward the Ph.D. degree in information and communication technologies.

He is with the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. His current research interests are in the areas of remote sensing, image classification, pattern recognition, and machine learning.

Mr. Persello is a Referee for the *Canadian Journal of Remote Sensing* and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Lorenzo Bruzzone (S'95–M'98–SM'03) received the Laurea (M.S.) degree in electronic engineering (*summa cum laude*) and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

From 1998 to 2000, he was a Postdoctoral Researcher with the University of Genoa. In 2000, he joined the University of Trento, Trento, Italy, where he is currently a Full Professor of telecommunications and the Head of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science. He teaches remote sensing, pattern recognition, radar, and electrical communications. His current research interests are in the areas of remote-sensing image processing and recognition (analysis of multitemporal data, feature extraction and selection, classification, regression and estimation, data fusion, and machine learning). He conducts and supervises research on these topics within the frameworks of several national and international projects. He is an Evaluator of project proposals for many different governments (including the European Commission) and scientific organizations. He is the author or coauthor of 74 scientific publications in referred international journals, more than 140 papers in conference proceedings, and 7 book chapters. He is a Referee for many international journals and has served on the Scientific Committees of several international conferences. He is a member of the Managing Committee of the National Inter-University Consortium for Telecommunications and of the Scientific Committee of the India–Italy Center for Advanced Research.

Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (Seattle, July 1998). He was a recipient of the Recognition of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Best Reviewers in 1999 and was a Guest Editor of a Special Issue of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING on the subject of the analysis of multitemporal remote-sensing images (November 2003). He was the General Chair and Cochair of the First and Second IEEE International Workshops on the Analysis of Multi-temporal Remote-Sensing Images (MultiTemp) and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003, he has been the Chair of The International Society for Optical Engineers Conference on Image and Signal Processing for Remote Sensing. From 2004 to 2006, he was an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and is currently an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. Since 2009, he has been a Member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society. He is also a member of the International Association for Pattern Recognition and the Italian Association for Remote Sensing (AIT).