Automatic Spectral Rule-Based Preliminary Classification of Radiometrically Calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye, and DMC/SPOT-1/-2 Imagery—Part II: Classification Accuracy Assessment

Andrea Baraldi, Laurent Durieux, Dario Simonetti, Giulia Conchedda, Francesco Holecz, and Palma Blonda, *Member, IEEE*

***Abstract*—In Part I of this paper, an operational fully automated Landsat-like image spectral rule-based decision-tree classifier (LSRC), suitable for mapping radiometrically calibrated seven-band Landsat-4/-5 Thematic Mapper (TM) and Landsat-7 Enhanced TM+ (ETM+) spaceborne images [eventually synthesized from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and the Moderate Resolution Imaging Spectroradiometer (MODIS) imaging sensor] into a discrete and finite set of spectral categories, has been downscaled to properly deal with spaceborne multispectral imaging sensors whose spectral resolution overlaps with, but is inferior to Landsat's, namely: 1) Satellite Pour l'Observation de la Terre (SPOT)-4/-5, Indian Remote Sensing Satellite (IRS)-1C/-1D/-P6 Linear Imaging Self-Scanner (LISS)-III, and IRS-P6 Advanced Wide Field Sensor (AWiFS); 2) National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) and Meteosat Second Generation (MSG); 3) Environmental Satellite (ENVISAT) Advanced Along-Track Scanning Radiometer (AATSR); 4) GeoEye-1, IKONOS-2, QuickBird-2, OrbView-3, TopSat, KOrean MultiPurpose SATellite (KOMPSAT)-2, FORMOsa SATellite (FORMOSAT)-2, Advanced Land Observing Satellite (ALOS) Advanced Visible and Near Infrared Radiometer type 2 (AVNIR-2), RapidEye, WorldView-2, PLEIADES-1/-2, and SPOT-6/-7; and 5) Disaster Monitoring Constellation (DMC), IRS-P6 LISS-IV, and SPOT-1/-2. LSRC, together with its five downscaled versions, identified, respectively, as the four-band SPOT-like SRC (SSRC), the four-band AVHRR-like SRC (AVSRC), the five-band AATSR-like SRC (AASRC), the four-band IKONOS-like SRC (ISRC), and the three-band DMC-like SRC (DSRC), form the so-called integrated SRC system of systems. In this paper, first, the classification accuracy and robustness to changes in the input data set of SSRC, AVSRC, AASRC, ISRC, and DSRC are assessed, both qualitatively and quantitatively, in comparison with LSRC's. Next, ongoing and future SRC applications are presented and discussed. They encompass: 1) the implementation of operational two-stage stratified hierarchical Remote Sensing (RS) image understanding systems discussed in Part I of this paper; 2) the integration of near real-time satellite mapping services with Internet map servers; and 3) the development of a new approach to semantic querying of large-scale multisensor image databases. These experimental results and application examples prove that the integrated SRC system of systems is operational, namely, it is effective, near real-time, automatic, and robust to changes in the input data set. Therefore, SRC appears eligible for use in operational satellite-based measurement systems such as those envisaged by the ongoing international Global Earth Observation System of Systems (GEOSS) Programme and the Global Monitoring for Environment and Security (GMES) system project.**

***Index Terms*—Decision-tree classifier, image classification, inductive and deductive inference, prior knowledge, radiometric calibration, remote sensing (RS).**

Manuscript received August 1, 2008; revised December 31, 2008 and July 1, 2009. First published December 4, 2009; current version published February 24, 2010.

A. Baraldi was with the Joint Research Centre, European Commission, 21020 Ispra, Italy. He is now with Baraldi Consultancy in Remote Sensing, 40129 Bologna, Italy (e-mail: andrea6311@gmail.com).

L. Durieux was with the Joint Research Centre, European Commission, 21020 Ispra, Italy. He is now with the Institut de Recherche pour le Développement, Maison de la Teledetection, 34093 Montpellier Cedex 05, France (e-mail: Laurent.durieux@ird.fr).

D. Simonetti is with the Global Environment Monitoring Unit, Institute for Environment and Sustainability, Joint Research Centre, European Commission, 21020 Ispra, Italy (e-mail: dario.simonetti@ext.jrc.it).

G. Conchedda is with the European Commission, B-1049 Bruxelles, Belgium (e-mail: giulia.conchedda@jrc.it).

F. Holecz is with Sarmap s.a., Cascine di Barico, 6989 Purasca, Switzerland (e-mail: fholecz@sarmap.ch).

P. Blonda is with the Instituto di Studi sui Sistemi Intelligenti per l'Automazione, Consiglio Nazionale delle Ricerche, 70126 Bari, Italy (e-mail: blonda@ba.issia.cnr.it).

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

Digital Object Identifier 10.1109/TGRS.2009.2032064

I. Introduction

FIVE downscaled implementations of the operational fully automated seven-band Landsat-like image spectral rule-based decision-tree classifier (LSRC), recently proposed to the remote sensing (RS) community [1], have been presented in Part I of this paper. These novel classifiers are identified as the four-band Satellite Pour l'Observation de la Terre (SPOT)-like SRC (SSRC), the four-band Advanced Very High Resolution Radiometer (AVHRR)-like SRC (AVSRC), the five-band Advanced Along-Track Scanning Radiometer (AATSR)-like SRC (AASRC), the four-band IKONOS-like SRC (ISRC)

and the three-band Disaster Monitoring Constellation (DMC)-like SRC (DSRC) (refer to Section III in Part I). In practice, the LSRC system together with its five downscaled implementations, hereafter identified as the integrated SRC system of systems, requires as input a multispectral (MS) image, radiometrically calibrated into top-of-atmosphere (TOA) reflectance (TOARF) values or surface reflectance ρ values (the latter being an ideal atmospheric noise-free case of the former, refer to Section II-B in Part I), acquired by almost any of the existing or future planned spaceborne Earth observation (EO) optical sensors. Thus, SRC appears eligible for use in operational satellite-based measurement systems such as those envisaged by the ongoing international Global EO System of Systems (GEOSS) Programme, conceived by the Group on EOs (GEO) [2], [3], and the Global Monitoring for Environment and Security (GMES) system project, an European Union initiative in partnership with the European Space Agency (ESA) [4].

To be consistent with the experimental session validity criteria proposed in [5],[1] the goal of this paper is to assess the classification accuracy and robustness to changes in the input data set of the novel downscaled SSRC, AASRC, AVSRC, ISRC, and DSRC systems in comparison with LSRCs in both synthetic and real-world RS image classification experiments.

The rest of this paper is organized as follows. In Section II, the classification accuracies of SSRC and LSRC are assessed and compared quantitatively in a real-world RS image classification problem at local scale where four alternative RS data sources, featuring different combinations of spectral and spatial resolution, are selected and compared in terms of information content and usability. In addition, to remove effects due to differences in spatial resolution and radiometric calibration, SSRC is compared quantitatively against LSRC in a synthesized testing data set. In Section III, ISRC is quantitatively and qualitatively assessed in one synthetic and two real-world data sets in comparison with LSRC. Examples of classification maps generated from AVSRC, AASRC, and DSRC at regional and continental scale are assessed qualitatively in Section IV. Finally, examples of ongoing and future applications of the operational integrated SRC system of systems are discussed in Section V. Conclusions of this experimental work are reported in Section VI.

II. Quantitative and Qualitative Performance Assessment of SSRC in Comparison With LSRC

The objective of this section is twofold. First, the SSRC classification accuracy and robustness to changes in the input data set are investigated quantitatively in one real-world and one synthetic RS image classification problem in comparison with LSRC's. Second, in agreement with the Quality Assurance Framework for EO (QA4EO) data, an international initiative led by the Committee of EOs (CEOS) Working Group on Calibration and Validation (WGCV) in the visionary framework of a GEOSS conceived by the GEO [2], [3], [7], [8], alternative RS data sources are compared in terms of metrological/statistically based radiometric and geometric quality indicators [8] required to automate the quantitative analysis of RS data (refer to Section I in Part I).

A. Test Case 1: Comparison of Multisource Classification Maps Generated From Ordinary One-Stage and Novel Two-Stage Classifiers

The following spaceborne image classification problem is considered representative of a large portion of RS image classification problems met in common practice where adequate reference (training/testing) data sets derived from field sites, existing maps and tabular data are expensive, difficult, or impossible to collect (refer to Section I in Part I).

In agreement with guidelines of the QA4EO initiative [8], in the present section, four RS image data sources featuring different combinations of spatial and spectral resolution and two alternative classification approaches are compared in the classification of local-scale mangrove ecosystems located in Western Senegal.

The four alternative RS image data sources selected for this experiment are the following: 1) a seven-band 30-m resolution Landsat-7 Enhanced Thematic Mapper Plus (ETM+) image; 2) a seven-band panchromatic (PAN)-sharpened 15-m Landsat-7 ETM+ image; 3) a four-band 10-m resolution SPOT-5 image; and 4) a seven-band Landsat-like 30-m resolution Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) image.

The two competing classifiers are a reference standard one-nearest-neighbor classifier, identified as 1-NN, and its stratified version, identified as 1-SNN, consisting of an operational automatic learning-by-rule preliminary SRC first stage in series with the supervised learning-by-example 1-NN classifier. Thus, 1-SNN represents an instantiation of the Shackelford and Davis two-stage stratified hierarchical RS image understanding system (RS-IUS) architecture presented and discussed in Section II-C3 of Part I. From a theoretical standpoint, it is known that, "stratification will always achieve greater precision provided that the strata have been chosen so that members of the same stratum are as similar as possible in respect of the characteristic of interest" [9]. In this experimental session, if 1-SNN outperforms 1-NN, then the effectiveness and robustness to changes in the input data set of the 1-SNN first stage, provided by SRC, are endorsed.

Finally, it is noteworthy that in the comparison between the 1-NN and 1-SNN classifiers, relative rather than absolute classification performance are of current interest.

1) Study Area: A test case was located across the downstream part and estuary of the river Casamance in southwestern Senegal. Known as Low Casamance, this study area covers a local geographic extent (areas up to 100 000 km^2 [24]) of about 5000 square km (70 $\times$ 70 km) defined by geographic coordinates between $12^\circ 10'$ and $14^\circ 10'$ latitude north and between $16^\circ 03'$ and $16^\circ 50'$ longitude west.

[1]According to [5], to provide a quantitative validation of a novel RS image classification approach, at least one real and one standard image classification problem, a battery of measures of success capable of dealing with the well-known noninjective property of any quality index [6], and at least one competing existing (well-known) data mapping algorithm (e.g., in our case, LSRC) ought to be selected for comparison purposes.

This study area is characterized by the presence of a multitude of mangrove-fringed channels and swamps interspersed between the tidal flats [10]–[13].

Behind and within *mangrove stands* are found the *salt mud flats*. These are bare and daily flooded areas which constitute a typical feature of mangrove ecosystems.

Between the water channels, on slightly higher ground than the *salt mud flats,* flat areas locally called *tannes* occur [8]. These zones are only occasionally tide influenced. They are usually bare soil, but may be covered with a layer of herbaceous halophytes.

In Low Casamance, rainfed rice paddies typically occur on the same morphological features as tannes. Common herbaceous crops are millet, groundnut and sorghum. Cultivation is generally limited to the upper lands which are never flooded. These lands form what is called *plateau* in French literature. A plateau is typically occupied by settlements, forest and tree fruit crops.

2) Target Land Cover Classes: According to Congalton, the land cover class taxonomy must be [14]: 1) mutually exclusive, i.e., each mapped area falls into one and only one class and 2) totally exhaustive [15], [16]. To satisfy this second condition, outliers fall into class "other." It is noteworthy that the definition of a rejection rate is a well-known objective of any RS image classification system, e.g., refer to [17, p. 185]. Nonetheless, in RS common practice image classifiers are often applied without any outlier detection strategy.

In this paper, the totally exhaustive set of mutually exclusive land cover classes comprises eight classes defined as follows, in agreement with the land cover taxonomy proposed in [15] and [16].

1) Mangrove vegetation. It includes subclasses such as: healthy mangrove, degraded mangrove, highly degraded mangrove, dead mangrove, reforestation, and natural regeneration. Therefore, it is expected to feature a significant within-class spectral variance.
2) Tree cover with canopy cover above 60%. This class corresponds to open forest of palm trees interspersed with fruit trees such as cashew nuts and mangos and by many wild fruit trees largely used for commercial purposes (e.g., the *Detarium senegalensis).* Thus, rather than a land cover class, this tree cover is a land use class equivalent to a combination of multiple land cover classes.
3) Tree (woody) savanna. This class corresponds to the woody savanna and savanna classes as defined in [15] and [16]. It can be described as lands with herbaceous and other understory systems and with forest height ≥ 2 m and canopy cover between 10% and 60%. Due to its multiple land cover nature, it is not a "pure" land cover class. Rather, it can be considered a land use class.
4) Herbaceous crops. In [15] and [16], this is defined as lands covered with temporary crops followed by harvest and a bare soil period. In our experiment, this class consists of several land covers, either rice, sorghum, or groundnut. Like the tree cover class and the tree savanna class, the herbaceous crop type is a land use class equivalent to a combination of multiple land cover classes.
5) Tannes (see Section II-A1), which are occasionally flooded. They may be covered with a layer of herbaceous halophytes although they are usually bare soil.
6) Salt mud flats (see Section II-A1), which are bare soil and daily flooded areas.
7) Water, including sea, river, and channel water.
8) All the rest of the world, identified as "Outliers."

3) Fieldwork: During the dry (winter) season, from November to December 2006, a stratified field campaign was conducted across the study area. Field data were collected *in situ* by means of a hand-held palm device (Personal Digital Assistant or PDA) incorporating a Global Positioning System (GPS), namely, a Garmin iQUE 3600. Exploitation of Cybertracker software (www.cybertracker.co.za) allowed the development of a customised screen sequence. Unfortunately, the cardinality of the reference data set acquired *in situ* was limited by logistic problems. For example, due to difficulties in reaching mangrove swamps, only 150 reference points were collected for class mangrove.

4) Target Output Products: In agreement with the QA4EO guidelines [8], this experimental work requires as output products: 1) one classification map generated from every input RS image; 2) a battery of community-agreed classification accuracy measurements suitable for dealing with the well-known noninjectivity of any quality index [6]; and 3) a confidence interval associated with every quality measurement.

It is noteworthy that, although forgotten in common practice, any evaluation measure is inherently noninjective [6]. For example, in classification map accuracy assessment and comparison, different classification maps may produce the same confusion matrix and different confusion matrices may generate the same confusion matrix accuracy measure. These observations suggest that no single universally acceptable measure of accuracy, but instead a variety of quality indices should be employed in practice [14], [18].

a) Classification map requirements: No minimum mapping unit is required. This is tantamount to saying that the classification system should be able to detect small but genuine image details at the sensor resolution, i.e., at pixel size.

b) Classification accuracy measurements with a confidence interval: In spite of being regarded as common knowledge in the RS community, computation of the training and testing data set cardinality as a function of the target classification map accuracy and error tolerance is often ignored in RS common practice in disagreement with the new QA4EO guidelines [8]. To make this paper self-contained, this subject is discussed below.

There are many well-known measures of accuracy that can be derived from a (square) confusion matrix (error matrix), e.g., overall accuracy (OA) probability $p_{\mathrm{OA}} \in [0, 1]$, normalized accuracy, producer's accuracy, user's accuracy, KHAT (kappa) coefficient in $[-1, 1]$, KHAT variance, Z coefficient, etc. [6], [14], [18], [19].

It is well known that any classification OA probability estimate, $p_{\mathrm{OA}} \in [0, 1]$, is a random variable (sample statistic) with a confidence interval (error tolerance), $\pm\delta$, associated with it, where $0 < \delta < p_{\mathrm{OA}} \leq 1$. In other words $p_{\mathrm{OA}} \pm \delta$ is a function

of the specific testing data set used for its estimation, and vice versa [20]. For example, for a given reference testing sample set size M_{test} comprising independent and identically distributed reference samples (in practice, this hypothesis is violated in image mapping problems due to spatial autocorrelation between neighboring pixels) and an estimated classification accuracy probability p_{OA}, it is possible to prove that the half width δ of the error tolerance $\pm\delta$ at a desired confidence level (e.g., if confidence level $= 95\%$ then the critical value is 1.96) can be computed as follows [14], [20], [21]:

$$\delta = \sqrt{\frac{(1.96)^2 \cdot p_{\text{OA}} \cdot (1 - p_{\text{OA}})}{M_{\text{test}}}}. \tag{2.1}$$

For each cth class simultaneously involved in the classification process, with $c = 1, \ldots, C$, where C is the total number of classes, it is possible to prove that (refer to [14, p. 294])

$$\delta_c = \sqrt{\frac{\chi^2_{(1,1-\alpha/C)} \cdot p_{\text{OA},c} \cdot (1 - p_{\text{OA},c})}{m_c}}, \qquad c = 1, \ldots, C \tag{2.2}$$

where α is the desired level of significance, i.e., the risk that the actual error is larger than δ_c (e.g., $\alpha = 0.07$), $1 - \alpha/C$ is the level of confidence (e.g., if $\alpha = 0.07$ and $C = 7$, refer to Section II-A2, then $1 - 0.07/7 = 0.99$), and $\chi^2_{(1,1-\alpha/C)}$ is the upper $(1 - (\alpha/C)) * 100$th percentile of the chi-square distribution with one degree of freedom (e.g., if the level of confidence is $(1 - 0.07/7) = 0.99$, then $\chi^2_{(1,0.99)} = 6.63$).

If error tolerance $\pm\delta$ associated with the classification accuracy probability estimate p_{OA} is ignored [6], alternative classifiers can be ranked erroneously when differences in classification accuracy as small as a few percentage points occur (as a typical example found in RS literature not to be imitated, refer to [22]).[2]

In this experiment, target values and confidence intervals of community-agreed classification accuracy measures are selected as follows. The target one-class $p_{\text{OA}} \in [0, 1] \pm \delta = (2.1)$ is fixed at 0.85 $\pm$ 2%, in agreement with the U.S. Geological Survey (USGS) classification system constraints. The per-class classification accuracy, $p_{\text{OA},c} \in [0, 1] \pm \delta_{\text{c}} = (2.2)$, $c = 1, \ldots, C = 7$ (excluding outliers, refer to Section II-A2) should be about equal [23] and never below 70% [18], whereas a reasonable reference standard for δ_{c} is about 5%. Additional classification quality indices adopted hereafter are the class-specific producer's and user's accuracies and the KHAT (kappa) coefficient [14].

5) Temporal, Spatial, and Spectral RS Data Requirements: A list of application-specific quantitative RS data functional constraints was defined as follows.

1) *Spatial resolution.* A common rule of thumb is to select a spatial resolution of the RS optical sensor equal to 1/3 ÷ 1/2 of the size of the smallest detail of interest in the depicted scene [24]. In this problem, the spatial resolution is fixed at ≤ 40 m to detect the most significant mangrove-fringed channels interspersed between the tidal flats (see Section II-A2).
2) *Spectral resolution.* To facilitate vegetation-type detection (see Section V in Part I of this paper), at least one visible channel, either blue (B), green (G), or red (R), one near-infrared (NIR) channel and one medium infrared (MIR) channel are recommended. This spectral resolution requirement excludes from the possible RS data sources popular very high resolution (VHR) imaging sensors (e.g., IKONOS-2, QuickBird-2, etc.) and SPOT-1/-2 images, e.g., refer to Section III in Part I of this paper.
3) *Coregistration and geocoding.* Multitemporal RS image coregistration is mandatory in all classification and change detection techniques. A quantitative measure of coregistration quality is the root mean square (rms) of the Euclidean distance between the location of ground control points (GCPs) in image pairs. For example, due to the pixel-by-pixel nature of the change detection analysis, it is recommended that the rms error between any two dated images should not exceed 0.5 pixels [21]. In [25], a registration accuracy $< 1/5$ of a pixel is required to achieve a change detection error $< 10\%$. In this application example, with respect to a master image, the coregistration error of a slave image is required to be $\leq$ 1/2 of a pixel.
4) *Observation timing.* Based on seasonal considerations in tropical areas, the dry season, equivalent to the winter season, is recommended to reduce the risk of cloud cover. Historical RS data acquisitions should date no longer than ten years before the field campaign conducted in November–December 2006 (refer to Section II-A3).

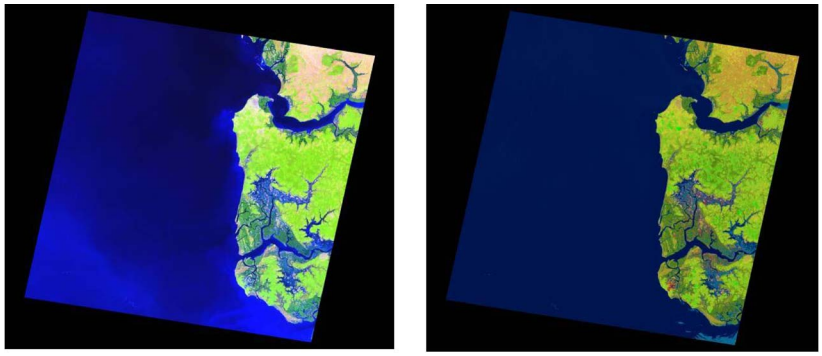
6) Selected Spaceborne Images: Three spaceborne optical images of the study area, free of clouds and capable of satisfying the temporal, spatial, and spectral RS data requirements listed in Section II-A5 were collected (see Table I).

1) One orthorectified georeferenced seven-band 30-m resolution Landsat-7 ETM+ image of the study area, acquired on November 6, 2000, was downloaded free of charge from the Global Land Cover Facility (GLCF) [26] [see Fig. 1(a)]. Thus, the time interval between the Landsat image acquisition and the ground truth observation campaign is approximately six years. This MS image is provided with a 15-m resolution PAN image.
2) One four-band 10-m resolution SPOT-5 image (level of processing L1B), acquired on March 8, 2006, was downloaded free of charge from the SPOT data archive in the

[2]For example, let the number of reference samples be $M = 100$. When this reference data set is input to a first classifier C_1 under testing, let the number of classification errors be $e_1 = 10$, then the estimated probability of error is $\hat{P}(e1) = 0.10$. If another classifier C_2 scores 13 misclassified testing samples in the same testing data set, then $\hat{P}(e2) = 0.13$. Unfortunately, at this point of the comparison between classifiers C_1 and C_2, it is impossible to state whether C_1 outperforms C_2. The confidence interval, identified as $\pm\delta$, at the 95% level of confidence of estimate $\hat{P}(e1)$ is as follows: $\hat{P}(e1) \pm \delta = \hat{P}(e1) \pm 1.96 \times \text{StDev}(\hat{P}(e1)) = 0.1 \pm 1.96 \times 0.030 = [0.04, 0.16]$ (due to the fact that if $M \geq 30$, and $e1$ is not close to zero, e.g., $e1 \geq 0.1$, a binomial distribution can be well approximated with a standardized normal distribution featuring zero mean and unit variance) [20]. A confidence interval at the 95% level of confidence of estimate $\hat{P}(e2)$ is as follows: $\hat{P}(e2) \pm \delta = \hat{P}(e2) \pm 1.96 \times \text{StDev}(\hat{P}(e2)) = 0.13 \pm 1.96 \times 0.036 = [0.06, 0.20]$. Since these two confidence intervals overlap, it is not possible to say that the classification accuracy of system C_1 will always be superior to that of system C_2.

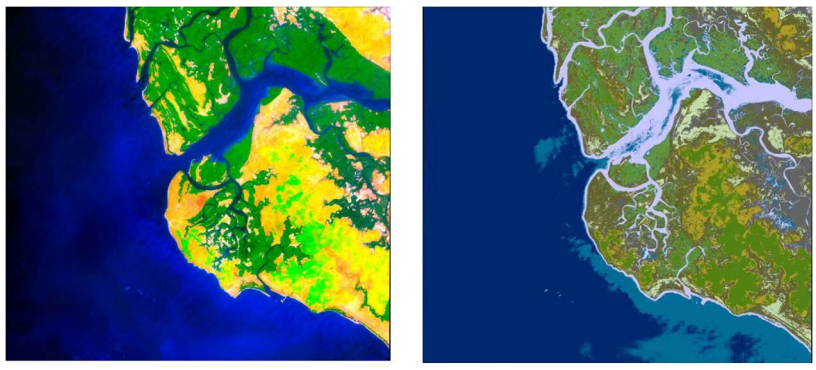
TABLE I
SYNOPSIS OF THE AVAILABLE SATELLITE DATA SET

LOW CASAMANCE target area			
Satellite sensor	**ASTER Terra**	**Landsat ETM+**	**SPOT-5**
Purchase strategy	**Purchased**	**Free access**	**Free access**
Date of acquisition	Feb. 28, 2004.	Nov. 06, 2000.	March 8, 2006.
Acquired scene(s)	2 (to be mosaicked: ASTL1B_040228113953060615 0077, ASTL1B_040228113953060615 0078)	1 (p205r051_20001106)	1 (SCENE 5 022-325 2006/03/08 11:50:54 1 J)
Spatial resolution (m)	G, R: 15m, NIR, SWIR*: 30 m, TIR: 90 m.	B, G, R, NIR, and SWIR*: 30 m, TIR: 60 m, PAN^: 15m.	G, R, NIR: 10 m, SWIR*: 20 m.
Number of available bands	14	7 (MS) + 1 (PAN)	4
Spectral range (μm)	0.52 - 11.65	0.45 - 12.50	0.50 - 1.75
Radiometrically calibrated testing image(s): bands, spatial resolution.	7 bands Landsat-like up- or down-sampled at 30 m resolution.	1) 7 bands (ETM1 to 5, ETM62, and ETM7) at 30 m resolution. 2) 7 bands PAN-sharpened at 15m resolution.	4 bands at 10 m resolution, where band SWIR was up-sampled by a factor of 2.
* SWIR=Short Wave Infrared = Medium IR (MIR), ^ PAN: Panchromatic.			



(a) (b)

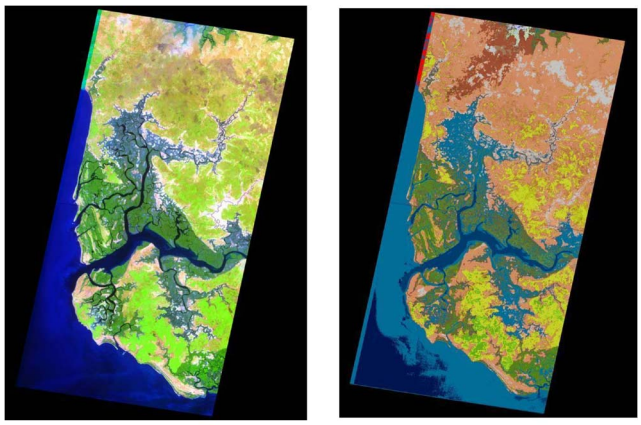
Fig. 1. (a) Landsat-7 ETM+ image of Senegal (path 205, row 051, acquisition date: November 6, 2000) in TOARF values, depicted in false colors (R: band ETM5, G: band ETM4, B: band ETM1), spatial resolution: 30 m. (b) Output map, depicted in pseudocolors, generated by LSRC from the radiometrically calibrated image shown in (a). Water and shadow areas are in blue, clouds in white, snow and ice in light blue, vegetation types in different shades of green, rangeland types in different shades of light green, and barren land types in different shades of brown and gray.



(a) (b)

Fig. 2. (a) SPOT-5 image of Senegal (path 325, row 022, acquisition date: March 8, 2006) in TOARF values and after dark-object subtraction, depicted in false colors (R: band 4, G: band 3, B: band 1), spatial resolution: 10 m. (b) Output map, depicted in pseudocolors [the same as in Fig. 1(b)], generated by SSRC from the radiometrically calibrated image shown in (a).

frame of the "Optimizing Access to SPOT Infrastructure for Science" program, funded by the European Commission for the European Scientific Community and operated by the Centre National d'Etudes Spaciales (CNES), France [see Fig. 2(a)]. The time interval between the SPOT image acquisition and the ground truth observation campaign is approximately 8 months. This image is affected by partial haze covering land masses.



(a) (b)

Fig. 3. (a) ASTER image mosaic of Senegal (acquisition date: February 28, 2004) in TOARF values, depicted in false colors (R: band 4, G: band 3, B: band 1), spatial resolution: 30 m. (b) Output map, depicted in pseudocolors [the same as in Fig. 1(b)], generated by LSRC from the radiometrically calibrated image shown in (a).

3) Two 14-band ASTER Terra images, acquired on February 28, 2004, were purchased at level 1B of processing [see Fig. 3(a)]. The time interval between the acquisition of the two ASTER images and the ground truth observation campaign is approximately two years and two months.

7) Image Preprocessing: The following image preprocessing steps were performed.

a) Image calibration into TOARF values, in agreement with the calibration/validation (Cal/Val) guidelines promoted by the QA4EO initiative [8].
   i. On the basis of existing literature [27]–[29], a sensor-specific software was developed to: 1) read the input image calibration metadata file; 2) transform digital numbers into TOA radiance (TOARD) values; and 3) transform TOARD into TOARF values (refer to Section II-B in Part I).
   ii. *Relative calibration.* In line with theoretical and operational considerations (refer to Section IV in Part I of this paper), an additional relative calibration step is required for the SPOT-5 image. In particular, by means of the Environment for Visualizing Images (ENVI) commercial software toolbox licensed by ITT Industries, Inc. [30], a dark-object subtraction operation was performed by matching the two spectral signatures of the same target water body identified in the radiometrically calibrated Landsat image, adopted as a master, and in the slave SPOT image.

b) Image band selection, combination, and resampling. Band 62 (thermal channel) of the Landsat image was upsampled from 60- to 30-m resolution with a cubic convolution (CC) approach provided by the ENVI software toolbox [30]. Band 4 of the SPOT image was upsampled from 20- to 10-m resolution with a CC approach. To generate a 30-m resolution seven-band Landsat-like image starting from a 14-band ASTER data set, the original ASTER image bands were selected, resampled at 30-m resolution, and combined according to Table II.

c) *Image coregistration.* The orthorectified Landsat image downloaded from GLCF was adopted as the master image. No coregistration procedure was required by the SPOT image. A geolocation displacement of the slave ASTER image pair with respect to the master Landsat image was assessed as < 90 m west (W) and < 30 m north (N). These values are consistent with those provided by the Earth Remote Sensing Data Analysis Center software tool, equal to 77 m W and 10.5 m N [31]. To reduce the spatial distortion due to the aforementioned linear displacement, together with additional error sources not dealt with by the ASTER L1B products (e.g., Earth rotation angle, Earth surface elevation values, etc.), a manual slave image-to-master image warp method was adopted. First, around 110 GCPs were selected throughout the target surface area in the ETM3 band of the reference image adopted as a guideline. Next, the same GCPs (termed conjugate points [32]) were localized upon the slave ASTER image. Finally, a second-order polynomial warp was applied to the slave image. A postwarp in-house automatic image correlation software toolbox estimated an average displacement error below 10 m, corresponding to 1/3 of a 30-m pixel size. This spatial distortion value satisfies the coregistration requirement ($\leq$ 1/2 of a pixel size) recommended in Section II-A5.

d) *Image mosaicking.* The two georeferenced ASTER scenes were mosaicked by means of the ENVI software toolbox.

e) MS image PAN-sharpening followed by a dark-object subtraction, implemented by means of the ENVI commercial software toolbox [30], to reduce spectral distortion, if any, in agreement with the QA4EO guidelines. In particular, the radiometrically calibrated 30-m resolution Landsat-7 ETM+ image was PAN-sharpened with the Gram-Schmidt PAN-sharpening algorithm, implemented in ENVI, which was adopted as a community-agreed reference standard. It is well known that the Landsat-7 ETM+ PAN sensor spectral response sensitivity is zero in the blue part of the spectrum, then it increases with wavelength being relatively low in green, significant in red, and optimal in infrared [33]. This means that any PAN-sharpened Landsat-7 image is affected by a systematic sensor-specific spectral distortion in addition to an accidental spectral distortion, if any, introduced by the PAN-sharpening algorithm at hand. To reduce this expected spectral distortion, a dark-object subtraction step (see comments above in this paper) was applied to the 15-m resolution PAN-sharpened Landsat image to match the spectral signature of water bodies in the PAN-sharpened (slave) image with the corresponding spectral signature in the original radiometrically calibrated (master) Landsat image.

To summarize, four testing RS images were generated from three RS data sources: 1) a 10-m resolution four-band SPOT-5 image; 2) a 30-m resolution seven-band Landsat-7 ETM+ image; 3) a PAN-sharpened 15-m resolution seven-band Landsat-7 ETM+ image; and 4) a 30-m resolution seven-band Landsat-like ASTER image.

TABLE II
GENERATION OF A LANDSAT-LIKE IMAGE FROM AN ASTER IMAGE

Reference ETM+ Band	ASTER band	Resampling Method
ETM1	Copy of Channel 1 (CH1)	-
ETM2	CH1	Nearest Neighbor (downsampling from 15 to 30 m).
ETM3	CH2	Nearest Neighbor (downsampling from 15 to 30 m).
ETM4	CH3	Nearest Neighbor (downsampling from 15 to 30 m).
ETM5	CH4	None (30 m resolution)
ETM62	MEAN (CH13, CH14)	Cubic Convolution (upsampling from 90 m to 30 m).
ETM7	MEAN (CH5, CH6, CH7, CH8)	None (30 m resolution)

8) Selected Classifiers: In general, land cover types are distinguishable based on a selection of image features such as (refer to Section II.C in Part I): 1) spectral properties, either pixel- or (2-D) segment-based, either chromatic or achromatic (brightness); 2) textural properties [34]; 3) morphological properties [35]; 4) geometric properties, e.g., compactness, rectangularity, etc. [36]; 5) either topological (e.g., adjacency, inclusion, etc.) or nontopological (e.g., distance, angle, etc.) spatial relationships between image objects; and 6) temporal properties.

In the spaceborne medium spatial resolution (in range 15 to 40 m, refer to Section I in Part I) RS images selected in Section II-A7, target land cover types (refer to Section II-A2) feature no salient texture, morphological or geometric property, but pixel-based (context-insensitive) chromatic and achromatic properties exclusively.

Considered as a benchmark classifier for its conceptual and computational simplicity and its effectiveness in many experimental environments, a nonparametric instance-based (also called memory- or kernel-based) k-nearest neighbor (k-NN) supervised classifier was chosen from existing literature [37]–[40]. Its sole free parameter, to be user-defined, is the number k of reference samples (independent of their class) to be found in a hypersphere centered around the input data vector which is assigned to the class having the largest number of representatives inside the hypersphere. The drawback of the k-NN classifier is that it requires (unlike, for example, prototype classifiers like the well-known plug-in maximum likelihood [38]) all reference samples to be retained in memory [37], [40]. This requirement may become a problem when the reference data set size is large, which is not the case in this paper where the small sample size problem is rather likely to occur [41]. In this experimental session, whose objective is a relative comparison of alterative classification results rather than the optimization of a given classification model, for the sake of simplicity k was set equal to one.

A 1-NN classifier implementation was selected from the NETLAB library [40], consisting of more than 150 Matlab functions and scripts based on the approach and techniques described in [37], which can be downloaded free of charge from the World Wide Web [42].

Alternative to a traditional 1-NN, a two-stage stratified 1-NN classifier, identified as 1-SNN, was implemented in line with the Shackelford and Davis two-stage stratified hierarchical RS-IUS architecture proposed in [43]–[46] (refer to Section II-C3 in Part I of this paper). The proposed instantiation of the two-stage 1-SNN classifier comprises:

1) A fully automated per-pixel preliminary SRC first stage (refer to Section I). In particular, SSRC and LSRC were employed to classify the four available data sets. SSRC was input with the radiometrically calibrated four-band SPOT-5 image while LSRC was input with the seven-band Landsat-7 ETM+ image, the PAN-sharpened seven-band Landsat-7 ETM+ image and the seven-band Landsat-like ASTER image (refer to Section II-A7).
2) A supervised 1-NN classification second stage that incorporates the "stratified" or "layered" approach. Table III shows the pixel-based relationship between each land cover class-specific reference sample set and spectral-based semiconcepts detected by the first-stage SRC. This image-derived relationship is capable of providing the land cover classification problem with class-specific spectral-based locational constraints (candidate areas, strata).

9) Reference Data Set: In this section, based on input data set dimensionality $D \geq 1$ and the classification system model complexity equivalent to the total number of the system free parameters $FP \geq 0$, a required number of reference training samples is estimated. Next, in agreement with the QA4EO initiative [8], starting from a target classification accuracy and error tolerance (refer to Section II-A4b), a required number of reference testing samples is estimated.

Heuristics commonly applied to both training and testing data sets are presented hereafter.

a) Required training data set cardinality: In recent years, enhanced spectral, temporal, and spatial resolutions of RS sensors have increased the number of detectable land cover classes. These developments have dramatically increased the size (cardinality) of the ground truth Regions of Interest (ROIs) required to be representative of the true class-conditional distributions. In this image classification scenario, the well-known *small training sample size problem* (also called *Hughes phenomenon* or *curse of dimensionality*) is likely to occur, which causes inductive learning systems to be affected by poor generalization capability [34], [37], [38], [41], [47].

To mitigate the Hughes phenomenon, heuristic rules have been traditionally employed in the training phase of supervised data learning classifiers. For example, let us assume that $\sum_{c=1}^{C} m_{\mathrm{train},c} = M_{\mathrm{train}}$, where $m_{\mathrm{train},c}$ is the number of independent training samples belonging to each class $c = 1, \ldots, C$. Empirical constraints applied to the size of the training data set are listed as follows.

1) $m_{\mathrm{train},c} \in \{5^{*}D, 100^{*}D\}$ [20], [23], [48], thus $M_{\mathrm{train1}} \approx C^{*}(5^{*}D \div 100^{*}D)$, where D identifies input

TABLE III
SPECTRAL OVERLAP BETWEEN REFERENCE SAMPLES AND PRELIMINARY SPECTRAL MAPS GENERATED BY THE INTEGRATED SRC SYSTEM OF CLASSIFIERS. A SEMANTIC-BASED CROSS-CHECK CAN BE PERFORMED FOR EACH AVAILABLE IMAGE-SPECIFIC REFERENCE DATA SET

		Satellite dataset											
		SPOT-5			PAN-sharpened Landsat-7 ETM+, 15 m			Landsat-7 ETM+, 30 m			ASTER		
		Adptd ref. samples, Tot. 1310.	Rejctd ref. samples, Tot. 81 (6.2%).	Prelim. spctrl types (in decreasing order of represent-tatives)	Adptd ref. samples, Tot. 1225.	Rejctd ref. samples, Tot. 74 (6.0%).	Prelim. spctrl types (in decreasing order of represent-tatives)	Adpt d ref. samples. Tot. 1239.	Rejctd ref. samples, Tot. 60 (4.8%).	Prelim. spctrl types (in decreasing order of represent-tatives)	Adptd ref. samples, Tot. 1595.	Rejctd ref. samples, Tot. 75 (4.7%).	Prelim. spctrl types (in decreasing order of represent-tatives)
Casamanceground truth dataset	Man-grove	265	23	AVLNIR: 130, WVLNIR: 129, WE: 6.	250	11	AVMNIR : 134, AVHNIR : 105, AVLNIR: 11	252	9	AVMNIR : 179, AVHNIR : 9, AVLNIR: 64	303	16	AVLNIR: 224, WE: 41, WVLNIR: 36, ASRLNIR: 2.
	Tree cover	171	5	AVMNIR: 123, AVLNIR: 38, ASRLNIR: 10	193	16	AVHNIR: 179, AVMNIR: 9, SVHNIR: 5.	209	0	AVHNIR: 134, SVHNIR: 54, AVMNIR: 21.	233	3	ASRMNIR : 102, AVMNIR : 96, AVHNIR : 24, ASRHNIR : 6, ASRLNIR : 5.
	Tree savan-na	188	25	WE: 130, TNCL: 45, ASRLNIR: 13.	84	6	ASRLNIR: 37, ASRMNIR: 19, SBBHTIR: 13, ABBHTIR: 8, AHR: 7.	85	5	ASRMNIR: 59, ABBHTIR: 10, ASRHNIR: 7, ASRLNIR: 7, SBBHTIR: 5, AHR: 3.	214	10	SBBHTIR: 173, ABBHTIR: 15, SBBLTIR: 11, ABBLTIR: 10, ASRLNIR: 5.
	Herba-ceous crop	190	6	TNCL: 108, WE: 52, DBB: 30.	212	9	AVMNIR: 83, AVLNIR : 50, ASRLNIR: 32, ASRMNIR: 19, WE: 12, WVLNIR: 8, AVHNIR: 5, ABBHTIR: 1, ASRHNIR: 1, BBBHTIR: 1.	212	9	AVLNIR : 82, AVMNIR: 54, ASRMNIR: 29, ASRLNIR: 20, WVLNIR: 11, WE: 7, TWA: 4, AVHNIR: 3, ABBLTIR: 2.	266	16	SBBHTIR: 254, SBBHTIR: 6, SBBLTIR: 6.
	Tannes	177	2	DBB: 115, WE: 45, TNCL: 17.	145	12	ABBHTIR: 53, SBBHTIR: 21, EmptySU: 14, DBBHTIR: 11, ABBHTIR: 12, TWA: 6, SBBHTIR: 4, WE: 3, SBBLTIR: 2, ABBLTIR: 2, ABBLTIR: 1, BBBHTIR: 1, BBBHTIR: 1, DBBHTIR: 1.	132	25	ABBHTIR: 37, BBBHTIR: 36, TWA: 25, EmptySU: 9, SBBHTIR: 7, BBBHTIR: 7, ABBHTIR: 6, WV: 3, SBBHTIR: 2.	199	14	ABBLTIR: 52, TWA: 39, ABBHTIR: 29, SBBHTIR: 28, ABBHTIR: 18, BBBHTIR: 11, SBBHTIR: 8, BBBHTIR: 6, ABBLTIR: 2.
	Salt mud flat	106	20	DBB: 63, TNCL: 40, TWASH: 3	124	20	TWASH: 80, SLWASH: 44.	132	12	TWA: 119, SLWASH: 13.	176	16	TWA: 144, DBBLTIR: 14, ABBLTIR: 11, ABBHTIR: 4, ABBHTIR: 3.
	Water	213	11	TNCL: 196, TWASH: 11, SLWASH: 6.	217	0	TWA: 104, SLWASH: 102, DPWASH: 11.	217	0	TWA: 151, SLWASH: 42, DPWASH: 23, SHV: 1.	204	0	TWA: 135, SLWASH: 63, DBBLTIR: 4, DBBHTIR: 1, EmptySU: 1.

space dimensionality [48]. For example, in a four-band SPOT-5 image (see Section II-A6) where $D = 4$, then $m_{\text{train},c} \in \{20, 400\}$, $c = 1, \ldots, C$. In the case of a seven-band Landsat and ASTER image (see Section II-A6) where $D = 7$, then $m_{\text{train},c} \in \{35, 700\}$, $c = 1, \ldots, C$.

2) $m_{\text{train},c} \geq 30 \div 50$, i.e., $M_{\text{train2}} \approx C^*(30 \div 50)$, so that according to a special case of the central limit theorem, the distribution of many sample statistics becomes approximately normal, which is a basic assumption employed by several traditional classifiers [14], [17].
3) To avoid poor generalization capability of an induced classifier related to model complexity, the minimum number of per-class representative samples should be proportional to the number of system free parameters FP to be learned from training samples. For example, an approximate worst case limit on generalization is that correct classification of a fraction $(1 - \varepsilon)$ of new examples requires a number of training patterns at least equal to $M_{\text{train3}} \approx FP/\varepsilon$, where ε is the classification error percentage in [0, 1]. For example, in the case of a supervised k-NN classifier (refer to Section II-A8), the number of free parameters is $FP = 1$. Thus, if $\varepsilon = 0.15$ (equivalent to an OA of 85%, see above in this paper), then the number of training patterns required by the k-NN classifier is $M_{\text{train3}} \geq FP/\varepsilon \approx 10$ [49].

To conclude, $m_{\text{train},c} = \max\{m_{\text{train1},c}, m_{\text{train2},c}, m_{\text{train3},c}\}$, $c = 1, \ldots, C$, and $M_{\text{train}} = \max\{M_{\text{train1}}, M_{\text{train2}}, M_{\text{train3}}\}$.

b) Required testing data set cardinality: If a holdout resampling method is adopted for the assessment of the generalization capability of competing classifiers where typically 2/3 of the available labeled data set should be used for training and the remaining 1/3 for testing [52], then the reference sample set size, M_{holdout}, becomes $M_{\text{holdout}}(M_{\text{train}}) = M_{\text{test}} + M_{\text{train}} = (3/2) * M_{\text{train}}$, where $M_{\text{train}} = \max\{M_{\text{train1}}, M_{\text{train2}}, M_{\text{train3}}\}$ (see Section II-A9a).

Starting from a confidence interval (error tolerance) associated with a classification OA probability estimate, $p_{\text{OA}} \in [0, 1]$ [see (2.1)], M_{test} can be estimated as [14], [20]

$$M_{\text{test}} = \frac{(1.96)^2 \cdot p_{\text{OA}} \cdot (1 - p_{\text{OA}})}{\delta^2}. \tag{2.3}$$

In this experiment, where the target quality index pair $p_{\text{OA}} = 0.85 \pm \delta = 2\%$ at a 95% confidence level (see Section II-A4b), then $M_{\text{test}} \approx 1225$.

Starting from (2.2), for each cth class simultaneously involved in the classification process, with $c = 1, \ldots, C$, it is possible to prove that (refer to [14, p. 294])

$$m_{\text{test},c} = \frac{\chi^2_{(1,1-\alpha/C)} \cdot p_{\text{OA},c} \cdot (1 - p_{\text{OA},c})}{\delta_c^2}, \quad c = 1, \ldots, C. \tag{2.4}$$

For example, if the desired level of significance, i.e., the risk that the actual error is larger than δ_i, is $\alpha = 0.07$ and $C = 7$ (refer to Section II-A2), then the level of confidence $(1 - 0.07/7) = 0.99$, and the upper $(1 - (\alpha/C)) * 100$th percentile of the chi-square distribution with one degree of freedom becomes $\chi^2_{(1,0.99)} = 6.63$. In this experiment, where δ_c is fixed equal to 5% and $p_{\text{OA},c} = 85\%$ (refer to Section II-A4b), then $m_{\text{test},c} \approx 340$.

To summarize, if $m_{\text{test},c} \approx 340$, $c = 1, \ldots, 7$, with a per-class classification accuracy $p_{\text{OA},c} = 0.85$, $c = 1, \ldots, 7$, then the confidence interval (error tolerance) of the accuracy assessment $p_{\text{OA},c}$ at the confidence level equal to 99% is $\delta_c = 5\%$, i.e., $p_{\text{OA},c} \pm \delta_c = 0.85 \pm 0.0425 = [0.8075, 0.8925]$.

c) Reference data set augmentation: Unfortunately, the cardinality of reference samples collected *in situ* is far below requirements presented above in this paper. For example, only 150 reference points were collected *in situ* for class mangrove, below requirements of $m_{\text{train},c}$ (e.g., if the number of bands $D = 4$, then $m_{\text{train},c} \in \{30, 400\}$; if $D = 7$, then $m_{\text{train},c} \in \{35, 700\}$) and $m_{\text{test},c}$ (e.g., if $c = 1, \ldots, 7$, then $m_{\text{test},c} \approx 340$) (refer to this paper above). Thus, the reference data set was augmented by visual selection of target land cover class samples in high-resolution images of the study area provided by commercial geobrowsers (namely, Google Earth, 2006). After augmentation, approximately 300 reference samples were collected for class mangrove (see Table III). However, due to potential spatial misplacements and temporal differences between ground truth samples and spaceborne imagery, cleaning of erroneous reference samples had to be conducted per input image (refer to Section II-A9d).

d) Reference data set refinement using semantics: Inconsistencies between spaceborne imagery and reference ground truth samples can be due to: 1) land cover changes occurring in the time interval (e.g., multiannual in these experiments) between the ground truth data acquisition campaign and the RS image acquisition date; 2) accidental problems occurred in the acquisition/storage/recovery of ground truth samples; and/or 3) spatial inconsistencies between the orthorectification/coregistration of reference samples with spaceborne imagery.

In this paper, for each of the four input RS images radiometrically calibrated into TOARF values (refer to Section II-A7), the automatic SRC system of systems was run (see Figs. 1–3). Next, for each reference pixel, its semantic meaning was compared against the symbolic meaning of the conjugate point in the SRC map domain. Each reference sample affected by a semantic inconsistency with the SRC map (e.g., a reference sample belongs to the ground truth class, for example, *Tree Savanna*, refer to Section II-A2, while its preliminary spectral category index is, for example, either *water or shadow*, refer to Section II-D in Part I of this paper) was double checked by a domain expert to be either rejected or confirmed.

Table III shows results of the semantic cross-checking between the reference data set and the SRC map automatically generated from every testing image. These results reveal that the average rejection rate of reference samples across the four input images is about 5.4%. This rejection rate is statistically relevant and, in our opinion, representative of a potential source of uncertainty of classification accuracy measurements largely underestimated in RS common practice [50], [51]. It also means that the cleaning of reference ground truth samples per RS image based on semantics, which is made possible by the availability of the automatic SRC system of systems, is of potential interest to a large segment of the RS community.

e) Resampling method: If a typical holdout resampling strategy for a binary (e.g., mangrove versus nonmangrove) classification accuracy assessment is adopted (where 1/3 of the reference samples are used for testing and 2/3 for training [52]), since $m_{\text{holdout,mangrove}} \approx 300$ (refer to Section II-A9c), then $m_{\text{train,mangrove}} = 200$ and $m_{\text{test,mangrove}} = 100$. The training data set cardinality $m_{\text{train,mangrove}} = 200$ means that, if $D = 7$ (e.g., in the Landsat case), then $m_{\text{train,mangrove}} = 200 \approx 30 * (D = 7)$, which satisfies the three empirical criteria adopted in Section II-A9a to assess the minimum size of the training data set, $m_{\text{train},c} = \max\{m_{\text{train1},c}, m_{\text{train2},c}, m_{\text{train3},c}\}$, $c = 1, \ldots, C$. The testing data set cardinality $m_{\text{test,mangrove}} = 100$ means that, according to (2.4), an expected accuracy $p_{\text{OA,mangrove}} = 85\%$ at the confidence level equal to 99% features an error tolerance $\pm\delta$ as large as $\pm 9\%$, which is far superior to the uncertainty in accuracy measurement required in Section II-A4b.

To reduce the error tolerance caused by the holdout resampling strategy when the luxury of keeping aside part of the finite reference data set for estimating the generalization capability of a classifier cannot be afforded, N-fold cross-validation is recommended where the finite reference data set is randomly split into N parts, $(N - 1)$ parts for training

TABLE IV
SEVEN-BAND LANDSAT-7 ETM+ DATA SET. 1-NN CLASSIFICATION

		Reference class							
		Herbaceous	Mangrove	Mud flat	Tannes	Tree / Woodland	Tree savannah	Water	Sum
Predicted class	Herbaceous	181	1	3	1	7	12	0	205
	Mangrove	2	251	0	0	2	0	0	255
	Mud flat	5	0	108	7	0	0	31	151
	Tannes	3	0	4	123	0	1	0	131
	Tree / Woodland	4	0	0	0	200	1	0	205
	Tree savanna	17	0	0	1	0	71	0	89
	Water	0	0	17	0	0	0	186	203
	Sum	212	252	132	132	209	85	217	1239
	Producer's accu.	0,85	1,00	0,82	0,93	0,96	0,84	0,86	
	User's accu.	0,88	0,98	0,72	0,94	0,98	0,80	0,92	
	OA ± error intrvl	0,904 := (1120 / 1239) ± 1.64%							
	OA Stdev	0,003511885 := Stdev (0,911 0,907 0,904)							

TABLE V
SEVEN-BAND LANDSAT-7 ETM+ DATA SET. 1-SNN CLASSIFICATION

		Reference class							
		Herbaceous	Mangrove	Mud flat	Tannes	Tree / Woodland	Tree savannah	Water	Sum
Predicted class	Herbaceous	196	1	2	1	5	10	0	215
	Mangrove	1	251	0	0	1	0	0	253
	Mud flat	3	0	113	5	0	0	28	149
	Tannes	1	0	2	126	0	0	0	129
	Tree / Woodland	2	0	0	0	203	1	0	206
	Tree savanna	9	0	0	0	0	74	0	83
	Water	0	0	15	0	0	0	189	204
	Sum	212	252	132	132	209	85	217	1239
	Producer's accu.	0,92	1,00	0,86	0,95	0,97	0,87	0,87	
	User's accu.	0,91	0,99	0,76	0,98	0,99	0,89	0,93	
	OA ± error intrvl	0,930 := (1152 / 1239) ± 1.42%							
	OA Stdev	0,005196152 := Stdev (0,925 0,925 0,916)							

one specific classifier implementation and one part for testing the trained classifier [41]. If N is set equal to three, then three testing accuracy estimates can be combined (added). In this case, for class $mangrove$, the overall number of testing samples becomes $m_{\text{test,mangrove}} \approx 300$. As a consequence, an expected accuracy $p_{\text{OA,mangrove}} = 85\%$ at the confidence level equal to 99% features an error tolerance $\pm\delta_{\text{mangrove}} = (2.2) = \pm 5.3\%$, more in line with the target error tolerance fixed in Section II-A4b. Analogously to testing accuracy estimate, a threefold cross-validation estimate of training accuracy should be performed. However, the training error in a kernel-based NN classifier, such as the one selected in Section II-A8, is equal to zero (i.e., the NN classifier is also called a *consistent classifier* [53]) and can be ignored.

10) Classification Result Collection: For each of the four testing images (see Section II-A7) and for each of the two available classification methods, namely, a 1-NN classifier and its stratified version 1-SNN (see Section II-A8), the latter employing class-specific spectral strata selected according to Table III, and for each fold of a threefold cross-validation method (refer to Section II-A9e), one classifier was instantiated (trained), tested and a onefold classification confusion matrix was computed. Overall, 24 classifiers were implemented and tested and 24 onefold classification confusion matrices were computed. Next, per input image and per classification approach a threefold confusion matrix was generated by summing the three onefold confusion matrices. Overall, eight threefold confusion matrices were computed (one for each of the four input data sets times the two alternative classifiers), as shown in Tables IV–XI.

11) Discussion: Several general conclusions of potential interest to the RS community involved with the development of operational satellite-based measurement systems, such as those envisaged by the ongoing GEOSS and GMES programs, can stem from this application example where classification results are compared across sensors and classifiers.

TABLE VI
SEVEN-BAND PAN-SHARPENED LANDSAT-7 ETM+ DATA SET. 1-NN CLASSIFICATION

		Reference class							
		Herbaceous	Mangrove	Mud flat	Tannes	Tree / Woodland	Tree savannah	Water	Sum
Predicted class	Herbaceous	191	1	1	0	3	11	0	207
	Mangrove	2	249	0	0	2	0	0	253
	Mud flat	1	0	101	1	0	0	21	124
	Tannes	2	0	1	143	0	1	0	147
	Tree / Woodland	4	0	0	0	188	0	0	192
	Tree savanna	12	0	0	1	0	72	0	85
	Water	0	0	21	0	0	0	196	217
	Sum	212	250	124	145	193	84	217	1225
	Producer's accu.	0,90	1,00	0,81	0,99	0,97	0,86	0,90	
	User's accu.	0,92	0,98	0,81	0,97	0,98	0,85	0,90	
	OA ± error intrvl	0,931 := (1140 / 1225) ± 1.43%							
	OA Stdev	0,006110101 := Stdev (0,931 0,923 0,935)							

TABLE VII
SEVEN-BAND PAN-SHARPENED LANDSAT-7 ETM+ DATA SET. 1-SNN CLASSIFICATION

		Reference class							
		Herbaceous	Mangrove	Mud flat	Tannes	Tree / Woodland	Tree savannah	Water	Sum
Predicted class	Herbaceous	200	1	0	0	1	9	0	211
	Mangrove	1	249	0	0	1	0	0	251
	Mud flat	0	0	107	1	0	0	20	128
	Tannes	1	0	1	144	0	0	0	146
	Tree / Woodland	3	0	0	0	191	0	0	194
	Tree savanna	7	0	0	0	0	75	0	82
	Water	0	0	16	0	0	0	197	213
	Sum	212	250	124	145	193	84	217	1225
	Producer's accu.	0,94	1,00	0,86	0,99	0,99	0,89	0,91	
	User's accu.	0,95	0,99	0,84	0,99	0,98	0,91	0,92	
	OA ± error intrvl	0,949 := (1163 / 1225) ± 1.23%							
	OA Stdev	0,004582576 := Stdev (0,947 0,941 0,95)							

TABLE VIII
SEVEN-BAND LANDSAT-LIKE ASTER DATA SET. 1-NN CLASSIFICATION

		Reference class							
		Herbaceous	Mangrove	Mud flat	Tannes	Tree / Woodland	Tree savannah	Water	Sum
Predicted class	Herbaceous	206	0	1	13	0	41	0	261
	Mangrove	0	301	0	0	0	0	0	301
	Mud flat	0	2	149	18	0	0	4	173
	Tannes	11	0	23	164	1	1	0	200
	Tree / Woodland	0	0	0	1	230	3	0	234
	Tree savanna	49	0	1	2	2	169	0	223
	Water	0	0	2	1	0	0	200	203
	Sum	266	303	176	199	233	214	204	1595
	Producer's accu.	0,77	0,99	0,85	0,82	0,99	0,79	0,98	
	User's accu.	0,79	1,00	0,86	0,82	0,98	0,76	0,99	
	OA ± error intrvl	0,890 := (1419 / 1595) ± 1.53%							
	OA Stdev	0,012858201 := Stdev (0,899 0,895 0,875)							

TABLE IX
SEVEN-BAND LANDSAT-LIKE ASTER DATA SET. 1-SNN CLASSIFICATION

		Reference class							
		Herbaceous	Mangrove	Mud flat	Tannes	Tree / Woodland	Tree savannah	Water	Sum
Predicted class	Herbaceous	219	0	0	11	0	40	0	270
	Mangrove	0	303	0	0	0	0	0	303
	Mud flat	0	0	151	16	0	0	3	170
	Tannes	9	0	23	169	1	0	0	202
	Tree / Woodland	0	0	0	0	232	3	0	235
	Tree savanna	38	0	0	2	0	171	0	211
	Water	0	0	2	1	0	0	201	204
	Sum	266	303	176	199	233	214	204	1595
	Producer's accu.	0,82	1,00	0,86	0,85	1,00	0,80	0,99	
	User's accu.	0,81	1,00	0,89	0,84	0,99	0,81	0,99	
	OA ± error intrvl	0,907 := (1446 / 1595) ± 1.42%							
	OA Stdev	0,013316656 := Stdev (0,915 0,913 0,891)							

TABLE X
FOUR-BAND SPOT-5 DATA SET. 1-NN CLASSIFICATION

		Reference class							
		Herbaceous	Mangrove	Mud flat	Tannes	Tree / Woodland	Tree savannah	Water	Sum
Predicted class	Herbaceous	158	0	6	11	0	23	0	198
	Mangrove	0	253	0	0	0	0	1	254
	Mud flat	2	5	77	15	0	0	17	116
	Tannes	4	4	17	149	0	0	3	177
	Tree / Woodland	0	0	0	0	165	8	0	173
	Tree savanna	26	0	0	0	6	157	0	189
	Water	0	3	6	2	0	0	192	203
	Sum	190	265	106	177	171	188	213	1310
	Producer's accu.	0,83	0,95	0,73	0,84	0,96	0,84	0,90	
	User's accu.	0,80	1,00	0,66	0,84	0,95	0,83	0,95	
	OA ± error intrvl	0,879 := (1151 / 1310) ± 1.77%							
	OA Stdev	0,019756855 := Stdev (0,88 0,855 0,894)							

TABLE XI
FOUR-BAND SPOT-5 DATA SET. 1-SNN CLASSIFICATION

		Reference class							
		Herbaceous	Mangrove	Mud flat	Tannes	Tree / Woodland	Tree savannah	Water	Sum
Predicted class	Herbaceous	170	0	4	7	0	16	0	197
	Mangrove	0	263	0	0	0	0	0	263
	Mud flat	1	1	81	8	0	0	5	96
	Tannes	4	1	15	161	0	0	1	182
	Tree / Woodland	0	0	0	0	170	8	0	178
	Tree savanna	15	0	0	0	1	164	0	180
	Water	0	0	6	1	0	0	207	214
	Sum	190	265	106	177	171	188	213	1310
	Producer's accu.	0,89	0,99	0,76	0,91	0,99	0,87	0,97	
	User's accu.	0,86	1,00	0,84	0,88	0,96	0,91	0,97	
	OA ± error intrvl	0,928 := (1216 / 1310) ± 1.40%							
	OA Stdev	0,014843629 := Stdev (0,931 0,911 0,94)							

a) Comparison of RS imaging sensor quality factors: In agreement with the QA4EO guidelines, quality indexes associated with sensor derived data are of fundamental importance to automate the quantitative analysis of RS data (where "garbage in means garbage out"). In this experiment, the four available RS images are ranked as follows.

1) RS data dissemination and costs. Landsat-7 and SPOT-5 images were obtained free of charge. On November 20, 2008, the GEO announced that scientists and decision-makers around the world will soon have unrestricted access at no charge to the Landsat archive, the world's most extensive collection of continuously acquired remotely sensed satellite imagery [54].
2) Ease of use, inversely related to the manpower spent to guarantee the degree of image preprocessing required by an operational satellite-based measurement system to be developed in agreement with the QA4EO guidelines.
   a) Absolute calibration.
      i) Landsat and ASTER images were provided with a complete and reliable set of radiometric calibration parameters which guarantees radiometric calibration into TOARF values.
      ii) Affected by a lack of radiometric calibration offset parameters in the calibration metadata file, the available SPOT-5 image required a relative calibration stage, equivalent to dark-object subtraction (refer to Section II-A7), in series with absolute calibration. This real-world experience supports the general conclusion discussed in Section IV of Part I of this paper where the use of SPOT-5 imagery in an operational automatic satellite-based measurement system is not considered straightforward, requiring manual or semiautomatic data processing with a relevant degree of user supervision.
   b) Image band selection, combination, and resampling. Fast and simple operations were required by the Landsat and ASTER data sets (refer to Section II-A7).
   c) Image coregistration.
      i) The Landsat MS and PAN image pair were delivered orthorectified and geolocated.
      ii) SPOT-5 required no coregistration with the master Landsat image.
      iii) ASTER required coregistration with the master Landsat image.
   d) Image mosaicking. The two ASTER images were geolocated. As a consequence, their mosaic generation with a commercial software toolbox was straightforward.
   e) MS image PAN-sharpening.
      i) The 30-m resolution MS Landsat image was provided with a coregistered 15-m resolution PAN image free-of-charge. It is noteworthy that the Landsat MS and PAN image pair coregistration quality was quite satisfactory and no coregistration effort was undertaken before PAN-sharpening the MS image. The MS Landsat image PAN-sharpening procedure was implemented in a commercial image processing software toolbox; it was straightforward to run and performs well [55], although it required a dark-object subtraction stage to reduce spectral distortion with respect to the original MS Landsat image.
      ii) It is worth mentioning that the SPOT-5 MS 10-m resolution and PAN 2.5 ÷ 5-m resolution image pair features a time baseline around 3 s equivalent to a space baseline of about 25 km. This spatial displacement makes coregistration of a SPOT-5 MS and PAN image pair difficult and time consuming in RS common practice, particularly in images depicting terrain areas where it is difficult to detect conjugate point pairs (such as mountainous areas covered by snow). In RS common practice, the PAN-sharpening of an MS SPOT image can be considered very difficult to perform.

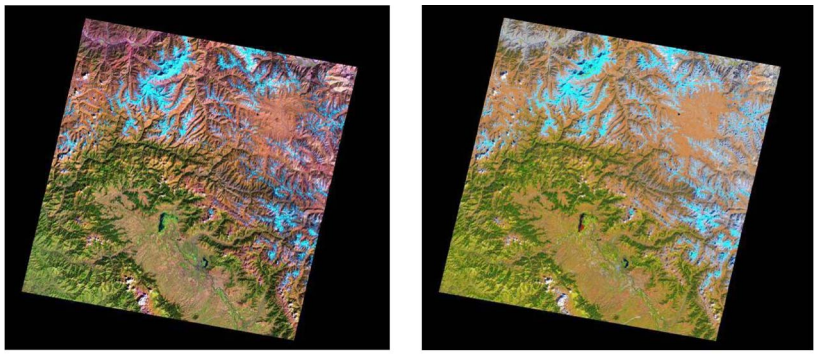
To summarize, in terms of cost, combination of spectral and spatial resolution, radiometric and geometric quality, dissemination and availability of PAN and MS imagery easy to fuse, the Landsat-7 mission appears superior to SPOT-5 and ASTER.

b) Classification accuracy results: Classification accuracy results collected in the comparison between 1-SNN and 1-NN across four input data sets are summarized below (refer to Tables IV–XI).

1) Thirty-meter resolution seven-band Landsat-7 ETM+, 1-NN: $p_{\mathrm{OA}} \pm \delta = 0.904 \pm 1.64\%$, 1-SNN: $p_{\mathrm{OA}} \pm \delta = 0.930 \pm 1.42\%$. Improvement due to first-stage color-based stratification: $+2.9\%$.
2) Fifteen-meter resolution seven-band PAN-sharpened Landsat-7 ETM+, 1-NN: $p_{\mathrm{OA}} \pm \delta = 0.931 \pm 1.43\%$, 1-SNN: $p_{\mathrm{OA}} \pm \delta = 0.949 \pm 1.80\%$. Improvement due to first-stage color-based stratification: $+1.9\%$.
3) Thirty-meter resolution seven-band Landsat-like ASTER, 1-NN: $p_{\mathrm{OA}} \pm \delta = 0.890 \pm 1.53\%$, 1-SNN: $p_{\mathrm{OA}} \pm \delta = 0.907 \pm 1.42\%$. Improvement due to first-stage color-based stratification: $+1.9\%$.
4) Ten-meter resolution four-band SPOT-5, 1-NN: $p_{\mathrm{OA}} \pm \delta = 0.879 \pm 1.77\%$, 1-SNN: $p_{\mathrm{OA}} \pm \delta = 0.928 \pm 1.40\%$, Improvement due to first-stage color-based stratification: $+5.5\%$.

These quantitative results show the following.

1) The two-stage 1-SNN classifier always achieved greater accuracy than the ordinary one-stage 1-NN classifier considered as a benchmark in existing literature. This is perfectly in line with the general principle of stratification quoted above in this paper (refer to the introduction to Section II-A) [9] and proves that preliminary spectral strata automatically detected by SRC are accurate symbolic image primitives robust to changes in the input data set acquired across time, space and sensors. This conclusion about the SRC effectiveness agrees with the first experimental evidence, collected in Section II-A9d, where semantic labels automatically provided by SRC allowed to reject an average 5.4% of the available reference data set across the four input images.
2) Both 1-SNN and 1-NN obtained their best classification performance with the 15-m resolution seven-band



(a) (b)

Fig. 4. (a) Landsat-7 ETM+ image in TOARF values of a rugged terrain area in Pakistan (path 149, row 036, acquisition date: September 30, 2001), depicted in false colors (R: band TM5, G: band TM4, B: band TM1). Starting from this Landsat image, a synthetic SPOT image is generated from a selection of the calibrated Landsat bands ETM2 to ETM5. (b) Output map generated by SSRC from the synthetic SPOT image extracted from the Landsat image shown in (a). In the output classification map, labeled pixels are depicted in pseudocolors [the same as in Fig. 1(b)].

PAN-sharpened Landsat-7 ETM+ image. This means that when the RS data source improves its spatial resolution but decreases its spectral resolution, e.g., in the comparison of a four-band 10-m SPOT-5 image with a seven-band PAN-sharpened 15-m Landsat-7 image, then the two noncontextual (pixel-based) image classifiers, 1-SNN and 1-NN, decrease their mapping accuracy. Vice versa, in line with theoretical expectations, when the RS data source enhances its spatial resolution with no loss in spectral resolution, e.g., in the comparison of PAN-sharpened 15-m Landsat-7 with 30-m Landsat-7 imagery, an improvement in the classification performance of both 1-SNN and 1-NN occurs.

Three conclusions stem from this experimental evidence. First, in agreement with existing literature, which considers Landsat-5 TM and Landsat-7 ETM+ data capable of having tremendous scientific utility and which urgently requires a follow-on mission to Landsat-7 [24, p. 451], these authors rank the PAN and MS Landsat-7 ETM+ satellite image pair as the reference standard among spaceborne optical data sources for land cover and land cover change product generation. As a consequence, these authors expect the universal availability of cost-free Landsat images announced by GEO to be capable of revolutionizing the use of EO data by scientists and decision-makers around the world [54]. Second, in line with theory (refer to Section II-C3 in Part I of this paper), this experimental section proves that a two-stage stratified hierarchical RS-IUS architecture employing an operational automatic integrated SRC system of systems as its preliminary classification first stage is as follows: 1) superior to its nonstratified ordinary classification counterpart in terms of effectiveness and 2) robust to changes in the input data set acquired across time, space and sensors. Third, in the context of an operational near real-time two-stage stratified hierarchical RS-IUS implementation (refer to Section II-C3 in Part I of this paper) [43]–[46], the impact of an input MS imaging sensor featuring finer spatial resolution counterbalanced by a coarser spectral resolution should be carefully assessed in advance in terms of community-agreed operational performance measurements (refer to Section I in Part I). This final observation is somehow in contrast with the recent popularity of ongoing or future planned VHR spaceborne missions whose enhanced spatial resolution, but inferior spectral resolution with respect to traditional medium (e.g., Landsat) and low [e.g., Moderate Resolution Imaging Spectroradiometer (MODIS)] spatial resolution spaceborne optical sensors make the former less suitable for automatic image understanding (also refer to Sections I and V in Part I).

B. Test Case 2: Spot-Like Image Synthesized From a Calibrated Landsat Image

In Section II-A11b, classification accuracies gained by two different classifiers were always superior when the input RS image was a seven-band PAN-sharpened 15-m resolution Landsat-7 ETM+ image rather than a four-band 10-m resolution SPOT-5 image. In Section III of Part I, the loss in the total number of detected spectral categories affecting the downscaled SSRC system in comparison with LSRC, due to the coarser spectral resolution of SPOT-like images with respect to Landsat-like images irrespective of spatial resolution, was estimated at around 30%.

The aim of this experiment is to assess quantitatively the expected loss in spectral discrimination capability affecting the downscaled SSRC system in comparison with LSRC due to the inferior spectral resolution of the former irrespective of intersensor differences in radiometric calibration and spatial resolution. To reach this objective, a synthesized SPOT-like image (i.e., comprising a G, R, NIR, and MIR1 band combination, refer to Section III in Part I) is generated from a radiometrically calibrated Landsat-7 ETM+ image of a rugged terrain area of Pakistan whose ETM bands 1 (Visible Blue, B), 6 (TIR), and 7 (MIR2) are removed (also refer to Table I) [see Fig. 4(a)]. Second, the LSRC system is run on the reference Landsat image while the SSRC classifier is run on the synthetic SPOT image [see Fig. 4(b)].

Table XII shows the occurrence of 13 spectral metacategories (generated from semantic-based aggregation of the 46 spectral categories detected as leaves of the LSRC decision-tree) in the reference LSRC map. Table XIII shows the confusion matrix between the reference LSRC map and the testing SSRC map in the vegetation/nonvegetation dichotomous classification

TABLE XII
LABELED PIXEL POPULATION IN THE REFERENCE LSRC MAP GENERATED FROM A LANDSAT-7 ETM+ IMAGE OF PAKISTAN (PATH 149, ROW 036, ACQUISITION DATE: SEPTEMBER 30, 2001), WHERE THE 46 OUTPUT SPECTRAL CATEGORIES OF THE LSRC SYSTEM PROPOSED IN [1] ARE GROUPED, BASED ON THEIR SEMANTIC CONTENT, INTO 13 SUPERCATEGORIES

Spctrl category index	Spectral type name	Acronym	Reference pixels in the (master) Landsat map
1	Vegetation = Strong/Average/Weak Vegetation (SV/AV/WV)	V	4483552
2	Shadow Vegetation = Shadow Vegetation (SHV) or Shadow Rangeland (SHR)	SHV	2133398
3	Rangeland = Strong/Average Shrub Rangeland (SSR/ASR) or Strong/Average Herbaceous Rangeland (SHR/AHR) or Dark Rangeland (DR) or Wetland (WE)	R	7512247
4	Weak Rangeland	WR	2837000
5	Pit Bog or greenhouse	PB	21785
6	Barren land or Built up = Bright/Strong/Average/Dark Barren land or Built up (BBB/SBB/ABB/DBB)	BB	18981980
7	Shadow Barren land	SHB	41073
8	Water or Shadow = Turbid Water (TWA) or Turbid Water or Shadow Snow (TWASHSN) or Deep/Shallow Water or Shadow (DPWASH/SLWASH)	WASH	870334
9	Thick Cloud	TKCL	304184
10	Thin Cloud	TNCL	1936597
11	Snow = Snow (SN) or Ice or Snow (ICSN)	SN	1532637
12	Shadow Cloud	SHCL	451302
13	Unknown	UN	342412
		Total	70831761

TABLE XIII
CONFUSION MATRIX BETWEEN THE REFERENCE LSRC MAP AND THE TESTING SSRC MAP GENERATED FROM A SYNTHETIC FOUR-BAND SPOT-4/-5-LIKE IMAGE WHEN THE DICHOTOMOUS VEGETATION/NONVEGETATION CLASSIFICATION PROBLEM IS CONSIDERED. DESPITE THE INFERIOR SPECTRAL RESOLUTION OF THE SPOT IMAGERY WITH RESPECT TO THE LANDSAT IMAGERY, THERE IS ALMOST NO INFORMATION LOSS ($< 1.5\%$) IN THE ONE-CLASS SSRC MAP WITH RESPECT TO THE REFERENCE LSRC MAP

		Reference pixels in the master classification map (percent)	
		Veg.	**Non-veg.**
Test classification map	**Veg.**	16450396 (96.84)	491972 (0.91)
	Non-veg.	537586 (3.16)	53351807 (99.09)
	Total pixels (percent)	16987982 (100)	53843779 (100)
	Overall Accuracy = (69802203/70831761) = 98.5465% ± 0% **Kappa Coefficient** = 0.9601		

problem. In this one-class classification problem, an OA of 98.55% reveals that, despite its inferior spectral resolution, a four-band SPOT-4/-5-like sensor's capability of separating vegetation from nonvegetation classes is almost equivalent to that of the seven-band Landsat-7 ETM+ sensor. This is in line with theory as reported in Section V of Part I. Since the vegetation/nonvegetation dichotomous classification task is located at the first decision level in several RS data classification hierarchical systems such as the USGS classification hierarchy [23] and the Food and Agriculture Organization of the United Nations (FAO) Land Cover Classification System (LCCS) [57], this experimental result is important because it provides experimental evidence of the consistency of the vegetation spectral category detected by SSRC with that detected by LSRC whose reliability has been tested over a widely distributed set of locations and time periods (e.g., refer to Section II-A in addition to [1]).

At their finest symbolic granularity, LSRC and SSRC generate two classification maps featuring a different number of label types (46 versus 32, according to Section III in Part I of this paper). To assess the goodness of fit between two generic labeled maps, i.e., two discrete maps whose finite number of discrete labels may not be the same and whose order of label types is arbitrary, the so-called overlapping area matrix (OAM), which is, in general, nonsquare, is adopted from the existing literature. If an OAM is square, it is known as follows [6], [56].

1) When a square OAM is generated from a generic pair of discrete labeled maps, referred to as the reference and the testing partition, whose order of label types is arbitrary, a square OAM may have to be reshuffled to maximize the sum of the main diagonal matrix elements before estimating the probability of error.
2) In a reshuffled square OAM where the sum of the main diagonal elements is maximized, the sum of offdiagonal elements is proportional to the probability of error, i.e., to the fraction of wrongly assigned pixels in the testing partition with respect to the reference partition.
3) In a square OAM reshuffled to maximize the sum of the main diagonal elements, the same accuracy measures developed for the well-known *classification confusion (error) matrix* (which is square, by definition) can be employed (e.g., the user's and producer's accuracies, the kappa coefficient, etc.).

Table XIV shows the nonsquare OAM computed between the reference LSRC map and the testing SSRC map where a semantic-based aggregation of spectral categories reduced their number from, respectively, 46 to 13 and 32 to 12. As

TABLE XIV
(NONSQUARE) OAM BETWEEN THE REFERENCE LSRC MAP AND THE TESTING SSRC MAP GENERATED FROM A SYNTHETIC SPOT IMAGE [6], [56]. IN DARK GRAY SHADE, MATCHING CELLS WITH PERCENTAGE $\geq 3\%$

		Ground truth pixels in the reference classification map (13 label types, in percentage) generated from a Landsat image												
		V	SHV	R	WR	PB	BB	SHB	WASH	TKCL	TNCL	SN	SHCL	UN
Test classification map (12 label types, in percentage) generated from a synthetic SPOT image	V	99.79	50.70						1.12		0.04		0.29	4.47
	SHV	0.1	39.80				0.14	1.58	0.06		0.08		0.17	12.83
	R		1.27	98.86							0.2			1.71
	WR		4.55		86.97		1.78		0.14		0.11			2.80
	PB	0.09		0.04		88.69				0.01	0.02			2.90
	BB		0.40				93.87	3.97	0.35	0.10	8.23			8.60
	SHB						0.02	91.38						1.70
	WASH	0.06	3.25		0.03		0.23	1.17	81.39		1.45		94.62	22.38
	TKCL						1.11			99.87	0.10			0.18
	TNCL (including SHCL)	0.05		1.09	13.00	10.96	2.84		16.20	0.03	89.46	0.26		18.64
	SN					0.34	0.01		0.69		0.19	99.74	4.90	15.90
	UN		0.02					1.79	0.07		0.30		0.01	7.89

a consequence, no diagonal and offdiagonal matching cells can be identified, i.e., no OAM OA can be computed. Rather, the nonsquare OAM can be employed to verify the semantic consistency of the testing map with the reference map. In OAM, for each matching cell (i, j) different from zero, with row index $i \in \{1, T\}$, where T is the total number of discrete labels in the testing map, and column index $j \in \{1, R\}$, where R is the total number of discrete labels in the reference map such that, in our case, $T \neq R$, if a semantic-based inconsistency between classes i and j does exist, then check (e.g., based on either visual inspection or comparison with other information sources like VHR imagery found in a 3-D Earth viewer such as Google Earth) whether this semantic inconsistency is due to wrong classification of either the reference samples belonging to class j (which, therefore, should be revisited, e.g., relabeled, removed, etc.), or testing samples belonging to testing class i (which, therefore, should be revisited, e.g., relabeled, removed, etc.). Table XIV reveals that semantic inconsistencies between the reference and the testing map are related to class Thin Cloud (TNCL). In other words, mainly due to the lack of the thermal channel, the SSRC capability of detecting class TNCL becomes extremely poor, affected by both a commission error and an omission error (the latter due to the confusion between the spectral response of thin clouds and light-toned soil types). This experimental evidence is perfectly in line with theoretical expectations about SSRC in comparison with LSRC (refer to Section III in Part I of this paper).

III. QUANTITATIVE AND QUALITATIVE ASSESSMENT OF ISRC

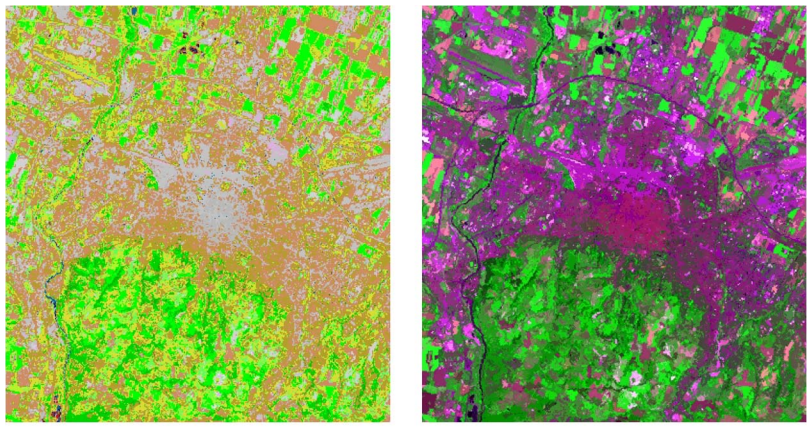
The quantitative assessment of the pixel-based automatic ISRC system is of potential interest to a large segment of RS community involved with the increasing popularity of VHR spaceborne images, the rapid growth of the commercial VHR satellite industry [73] and the development of object-based image analysis (OBIA) (refer to Section II-C2 in Part I), which is currently considered the state-of-the-art in VHR image understanding, alternative to traditional pixel-based image classifiers, despite OBIA is still affected by a lack of consensus on its conceptual foundation [58].

In this section, first, ISRC is quantitatively compared against LSRC in a test case employing as input a synthetic IKONOS-like image (i.e., comprising a B, G, R, and NIR band combination, refer to Section III in Part I) independent of intersensor differences in radiometric calibration and spatial resolution. Second, two ISRC preliminary classification maps, generated from a pair of real-world radiometrically calibrated IKONOS-2 and QuickBird-2 images, are visually assessed.

A. Test Case 1: IKONOS-Like Image Synthesized From a Calibrated Landsat Image

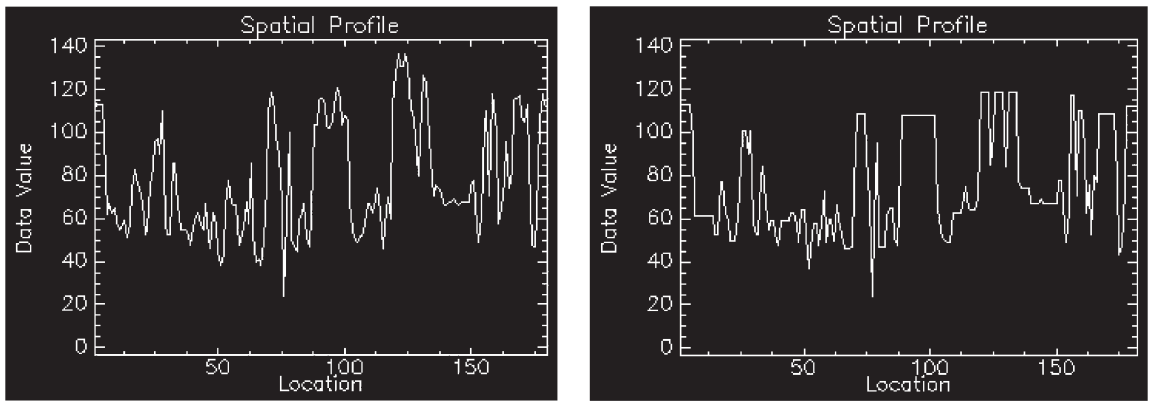
In Section III of Part I, the loss in the total number of detected spectral categories affecting the downscaled ISRC system in comparison with LSRC due to the coarser spectral resolution of four-band IKONOS-like images with respect to seven-band Landsat-like images irrespective of spatial resolution was estimated at about 45%.

The aim of this experiment is to assess quantitatively the expected loss in spectral discrimination capability affecting the downscaled ISRC system in comparison with LSRC irrespective of spatial resolution and intersensor differences in radiometric calibration. To reach this objective, a synthesized four-band IKONOS-like image is selected from a radiometrically calibrated seven-band Landsat-7 ETM+ image (path 192, row 029, acquisition date June 20, 2000; a spatial subset of this Landsat image is shown in Fig. 4(a) of Part I of this paper) where ETM bands 5 to 7 are removed (refer to Section III in Part I of this paper). A spatial subset of the output map automatically generated by ISRC from the synthesized four-band IKONOS-like image is shown in Fig. 5(a) (to be compared with Fig. 4(b) in Part I). A piecewise constant (low-pass, smoothed) approximation of the synthesized IKONOS-like input image is shown in Fig. 5(b). This piecewise constant image approximation is generated as follows: 1) extraction of symbolic segments from the ISRC map shown in Fig. 5(a), where a symbolic segment is defined as a connected area featuring



(a) (b)

Fig. 5. (a) Output map generated by ISRC from a synthetic IKONOS-like four-band image obtained from a radiometrically calibrated Landsat-7 ETM+ image of the city of Bologna, Italy, acquired on June 20, 2000, spatial resolution: 30 m, depicted in Part I of this paper as Fig. 4(a) (R: band 3, G: band 4, B: band 1), where ETM+ bands 5 to 7 are removed. Adopted pseudocolors are the same as in Fig. 1(b). (b) Piecewise constant approximation of Fig. 4(a) shown in Part I of this paper based on segments extracted from the preliminary classification map shown in (a), such that each segment is replaced with its mean reflectance value in the radiometrically calibrated image domain. It is noteworthy that while small but genuine image details appear well preserved, ISRC, despite being pixel-based, does not appear to be affected by the traditional salt-and-pepper classification noise effect. In practice, ISRC, which is capable of dealing with within-class variance, performs a context-(texture-)sensitive classification (e.g., see the image partition performed over urban areas).



(a) (b)

Fig. 6. (a) Transect extracted from band ETM4 of Fig. 4(a) in Part I. (b) Transect extracted from the piecewise constant approximation of the input band ETM4 provided by the image shown in Fig. 5(b). Small but genuine image details are well preserved. In practice, ISRC works as an edge-preserving smoothing filter while providing pixels with a semantic label.

the same label in the classification map domain (refer to Section II-C3b in Part I of this paper) and 2) replacement of each four-band pixel value in the IKONOS-like image with its segment-based mean reflectance value. When it is visually compared with the original Landsat image shown in Fig. 4(a) in Part I, the piecewise constant approximation of the IKONOS-like image shown in Fig. 5(b) reveals that genuine but small image details appear well preserved. This edge-preserving smoothing effect is clearly visible by comparing Fig. 6(a), depicting a transect extracted from band ETM4 of Fig. 4(a) in Part I, with Fig. 6(b) depicting the same transect extracted from band 4 of the piecewise constant IKONOS-like image approximation shown in Fig. 5(b). In practice, while mapping unlabeled pixels into a finite and discrete set of color-based semiconcepts, ISRC works analogously to an edge-preserving smoothing filter. This also means that, despite the fact it is pixel-based, ISRC does not appear to be affected by the well-known salt-and-pepper classification noise effect traditionally affecting pixel-based classifiers, i.e., ISRC successfully deals with within-class spectral variability (also refer to Section II.D in Part I of this paper). In practice, as shown by the classification performance of ISRC over urban areas which are highly textured, the capability of dealing with within-class variance allows ISRC to detect textons (texture elements) [59].

A visual comparison of the LSRC map shown in Fig. 4(b) in Part I with the ISRC map show in Fig. 5(a) reveals that the ISRC mapping capability of separating vegetation from nonvegetated surface types appears superior to that theoretically expected (refer to Section V in Part I). To provide this qualitative assessment with a quantitative estimate, an accuracy assessment of the ISRC map [shown in part in Fig. 5(a)] in comparison with the LSRC map (shown in part in Fig. 4(b) of Part I) is provided in Table XV. In this comparison, the LSRC map is called the master (reference) map and the ISRC map is the testing map. First, the 13 LSRC (reference) spectral supercategories listed in Table XII are reduced to eight metacategories by semantic-based aggregation. These semantic-driven combinations are described in the second column of Table XV.

TABLE XV
PRODUCER'S AND USER'S ACCURACIES, OMISSION AND COMMISSION ERRORS, OA, AND KAPPA ACCURACY COEFFICIENT EXTRACTED FROM A (SQUARE) CONFUSION MATRIX COMPUTED BETWEEN THE REFERENCE MAP GENERATED BY LSRC FROM A MASTER SEVEN-BAND LANDSAT IMAGE AND THE TESTING CLASSIFICATION MAP GENERATED BY ISRC FROM A SYNTHESIZED FOUR-BAND IKONOS-LIKE IMAGE. EIGHT CLASS SETS IN THE LEFT COLUMN ARE GENERATED BY SEMANTIC-BASED AGGREGATION OF THE 13 SUPERCATEGORIES LISTED IN TABLE XII

		Ground truth pixels in the reference classification map (percent)							
		V	SH	WR	BB	WASH	CL	SN	UN
Test classification map	V = V + R + PB	99.47	11.28		0.49				5.03
	SH = SHV + SHB	0.05	76.19	0.03	1.74	6.74	44.70		19.54
	WR								
	BB	0.37	2.03	99.94	84.79	6.91	3.03		47.97
	WASH		6.78		2.09	81.27	20.87		7.99
	CL = TNCL + TKCL + SHCL		0.32		0.10	2.43	31.39		1.05
	SN	0.02			10.72	0.09		100	15.68
	UN	0.10	3.41	0.03	0.07	2.54			2.74
	Total (percent)	100	100	100	100	100	100	100	100
	Producer's Accuracy (percent, pixels)	99.47% = 29146265/ 29301563	76.19% = 1717681/ 2254538	0% = 0/ 2451225	84.79% = 4947627/ 5835302	81.27% = 896019/ 1102483	31.39% = 5732/ 18258	100% = 248782/ 248782	2.74% = 7081/ 258379
	Omission (percent, pixels)	0.53% = 155298/ 29301563	23.81% = 536857/ 2254538	100% = 2451225/ 2451225	15.21% = 887675/ 5835302	18.73% = 206464/ 1102483	68.61% = 12526/ 18258	0% = 0/ 248782	97.26% = 251298 / 258379
	User's Accuracy (percent, pixels)	98.99% = 29146265/ 29442346	87.27% = 1717681/ 1968312	0% = 0/ 0	63.83% = 4947627/ 7751322	74.98% = 896019/ 1194944	11.86% = 5732/ 48325	27.01% = 248782/ 920914	4.90% = 7081/ 144367
	Commission (percent, pixels)	1.01% = 296081/ 29442346	12.73% = 250631/ 1968312	0% = 0/ 0	36.17% = 2803695/ 7751322	25.02% = 298925/ 1194944	88.14% = 42593/ 48325	72.99% = 672132/ 920914	95.10% = 137286 / 144367
	Overall Accuracy = 93.4405% = 64121516/68622859 ± 0% **Kappa Coefficient** = 0.8990								

Second, the same eight spectral categories are considered in the ISRC (testing) map. Thus, a square confusion matrix can be computed from the reference and testing map pair. Finally, a well-known set of error indexes (omission and commission errors) in combination with quality indexes (p_{OA} and kappa coefficient) are extracted from the standard (square) confusion matrix (see Table XV).

Starting from theoretical expectations (refer to Section V in Part I), it is quite surprisingly to observe that the p_{OA} of Table XV is above 90%, with a kappa coefficient around 0.90. In particular, the discrimination of spectral category *vegetation* (V) from the rest of the world is rather good (as it is affected by minor omission and commission errors) and somewhat better than expected. Some confusion problems exist between spectral category *either vegetation or bare soil in shadow*, SH (= SHV or SHB), and spectral category *either water or shadow*, WASH, which, by the way, are always difficult to discriminate. The main ISRC limitations affect spectral category *bare soil* or *built-up* (BB): several very bright BB pixels are classified as spectral category *snow* (SN). In addition, due to the loss in spectral resolution, spectral category *weak rangeland* (WR), featuring low leaf area index (LAI) values and consistent with the hypothesis of the presence of the land cover-type *herbaceous rangeland* (e.g., found in the USGS classification taxonomy), becomes, in general, indistinguishable from class BB (in fact, TOARF values of spectral categories BB and WR in the visible (VIS) and NIR portions of the electromagnetic spectrum appear identical in many practical cases).

To summarize, since the vegetation/nonvegetation dichotomous classification problem represents the first decision level in several RS data classification hierarchical systems such as the USGS classification hierarchy [23] and the FAO LCCS [57], Table XV is important because it provides experimental evidence of the consistency of the vegetation spectral category detected by ISRC with that detected by LSRC whose reliability has been tested over a widely distributed set of locations and time periods (e.g., refer to Section II-A in addition to [1]). In particular, Table XV shows that, if the spectral category WR (featuring low LAI values) is not considered, the separation of vegetation from the rest of the world is 99% consistent with that generated from LSRC, which is a result far better than that theoretically expected (refer to Section V in Part I of this paper).

In the near future, augmenting the ISPRC rule set over a large number of VHR test images may be a viable strategy in improving ISPRC overall classification performance. For example, new spectral rules can be conceived to detect man-made objects, such as the red tile roof of houses depicted at 1-m spatial resolution or finer, that are not visible at coarser spatial resolutions (refer to Fig. 12).

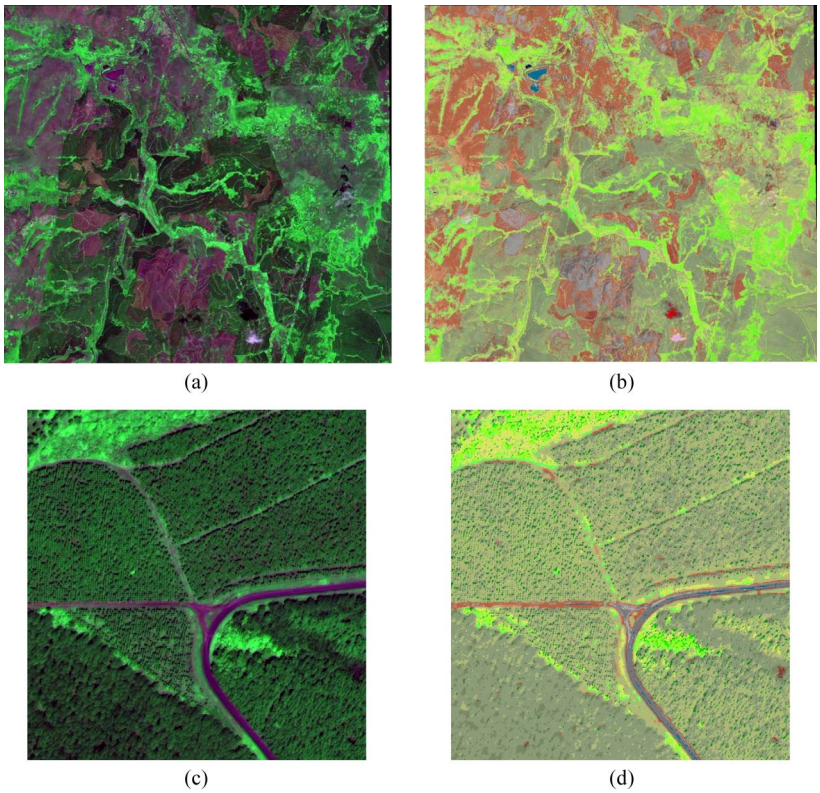


Fig. 7. (a) IKONOS-2 image of a forest and agricultural site in South Africa (acquisition date: November 14, 2006, 08:10 GMT) in TOARF values, depicted in false colors (R: band CH3, G: band CH4, B: band CH1), calibrated into TOARF values. Spatial resolution: 4 m. (b) Output map, depicted in pseudocolors [the same as in Fig. 1(b)], generated by ISRC from the radiometrically calibrated image shown in (a). (c) Zoomed image extracted from (a). Clearly visible are (moving from top to bottom): open forest and grassland, an unpaved road, a paved road, and several types of forest stand. (d) Zoomed map, extracted from (b), overlapping with (c).

B. Test Case 2: IKONOS-2 Image

An IKONOS-2 image (4-m spatial resolution, acquisition date: November 14, 2006 08:10, acquisition time 08:10 GMT) of a forest and agricultural area in South Africa is radiometrically calibrated into TOARF values [see Fig. 7(a)], and then classified by ISRC [see Fig. 7(b)]. To allow a qualitative visual assessment of the classification capability of ISRC, a local area extracted from Fig. 8(a) and (b) is shown in Fig. 8(c) and (d), respectively. In these figures, where a wide variety of forest stands is clearly visible, the classification capability of ISRC appears relevant and somewhat better than expected (refer to Section V in Part I of this paper).

C. Test Case 3: Quickbird-2 Image

A QuickBird-2 image of the Goro area in the Po river delta, Italy (2.4-m spatial resolution, acquisition date: May 28, 2003, time 12:45) is first radiometrically calibrated into TOARF values [see Fig. 8(a)], and then classified by ISRC [see Fig. 8(b)]. To allow a qualitative visual assessment of the ISRC mapping capability, a local area extracted from Fig. 9(a) and (b) is shown in Fig. 9(c) and (d), respectively. In these figures, where a wide variety of land cover types (e.g., urban areas, forest areas, sandy beaches, etc.) is clearly visible, the classification capability of ISRC appears impressive, somewhat better than expected (refer to Section V in Part I of this paper).

IV. Qualitative Assessment of AVSRC, AASRC, and DSRC

For the sake of completeness, Figs. 10–12 show preliminary classification maps generated by: 1) the AVSRC system running on a Meteosat second generation (MSG) image; 2) the AASRC system running on an AATSR image; and 3) the DSRC system running on a SPOT-2 image.

About the AVSRC preliminary spectral map at global scale shown in Fig. 9(b), generated from the radiometrically calibrated MSG image shown in Fig. 9(a), it is possible to observe that the former, comprising vegetation, bare soil, water, and cloud spectral types depicted in pseudocolors, appears consistent with a visual assessment of the latter. For example, this classification map shows vegetation spectral categories characterized by increasing greenness indexes (depicted in darker shades of green) at increasing distances from desert areas.

About the AASRC preliminary spectral map shown in Fig. 10(b) generated from the input AATSR image shown in Fig. 10(a), the former appears consistent with a visual assessment of the latter.

About the 20-m resolution SPOT-2 image shown in Fig. 11(a) after radiometric calibration into TOARF values, it should be compared with Fig. 2(a) showing a SPOT-5 image of the same area of Senegal at a spatial resolution of 10 m after radiometric calibration into TOARF values and dark-object subtraction (refer to Section II-A7). When DSRC is input with the RS

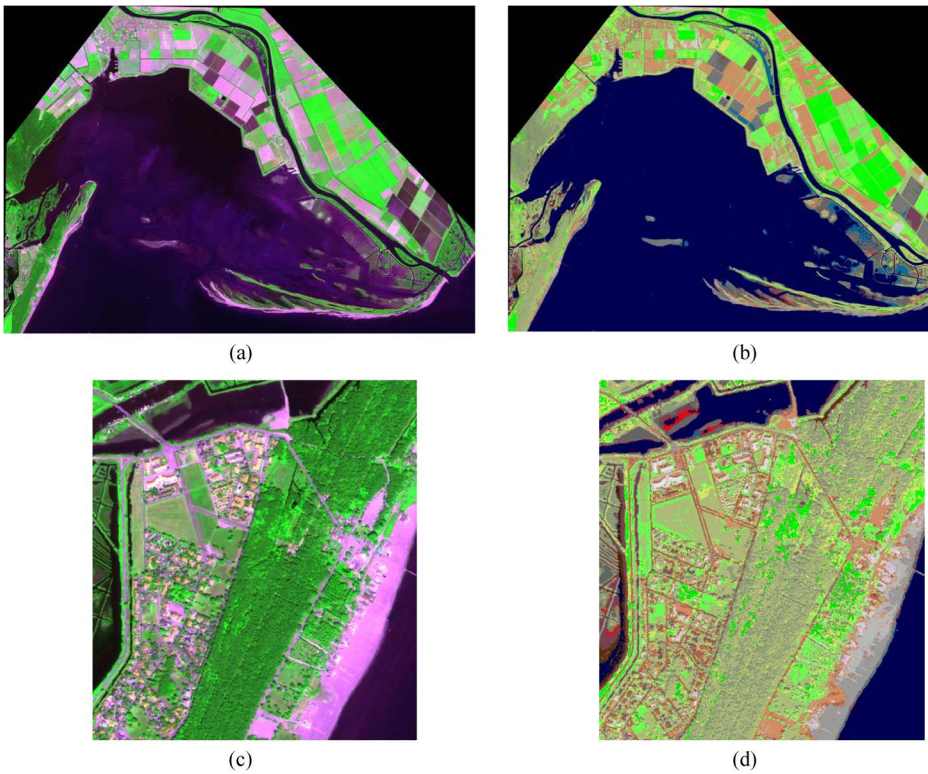


Fig. 8. (a) QuickBird-2 image of the Goro area in the Po river delta, Italy (acquisition date: May 28, 2003, 12:45), in TOARF values, depicted in false colors (R: band CH3, G: band CH4, B: band CH1), calibrated into TOA reflectance. Spatial resolution: 2.4 m. (b) Output map, depicted in pseudocolors [the same as in Fig. 1(b)], generated by ISRC from the radiometrically calibrated image shown in (a). (c) Zoomed image extracted from (a). Clearly visible are (moving from top to bottom): inner water, wetland, agricultural fields, urban areas, forest areas, sandy beach, and sea water. (d) Zoomed map, extracted from (b), overlapping with (c).

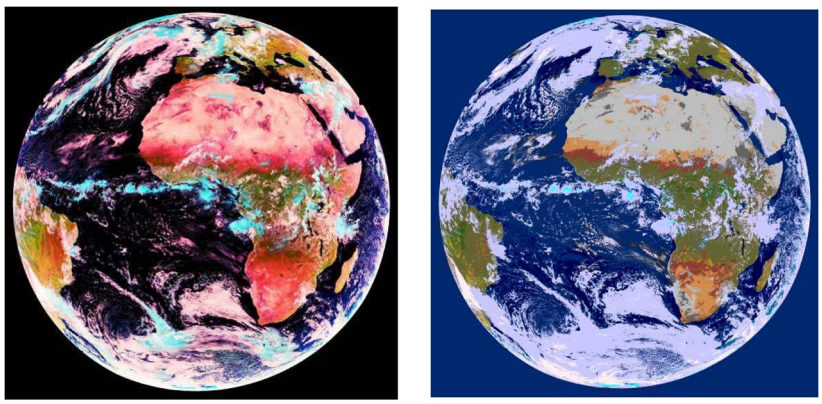
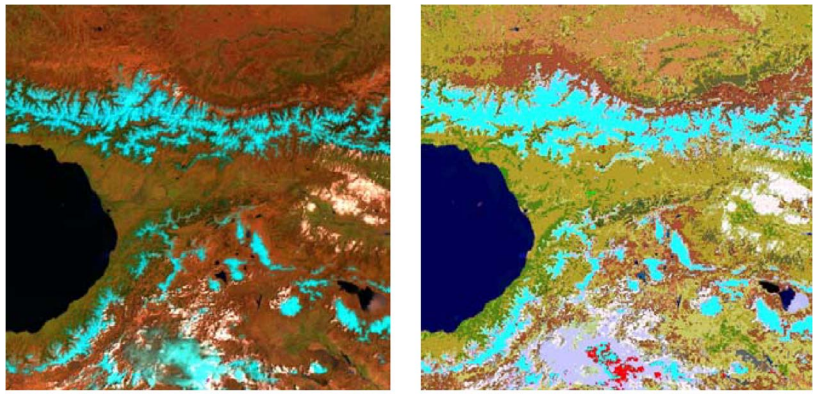


Fig. 9. (a) MSG image (acquisition date: May 16, 2007, 12:30) in TOARF values, depicted in false colors (R: band CH3, G: band CH2, B: band CH1). (b) Output map, depicted in pseudocolors [the same as in Fig. 1(b)] and featuring the large set of output categories, generated by AVSRC from the radiometrically calibrated image shown in (a).

image shown in Fig. 11(a), the preliminary classification map shown in Fig. 11(b) is generated, to be compared with the SSRC map shown in Fig. 2(b). Due to the LSRC large degree of redundancy, the downscaled DSRC version remains capable of generating a classification map consistent with a visual inspection of Fig. 11(a), although visible bands green (G) and red (R) feature a degree of interband correlation as large as 0.87 (whereas the correlation between band NIR and, respectively, band G and band R is 0.29 and 0.56). In practice, DSRC appears to be capable of performing the automatic classification of a radiometrically calibrated MS image featuring only two bands, namely, one visible and one NIR channel.

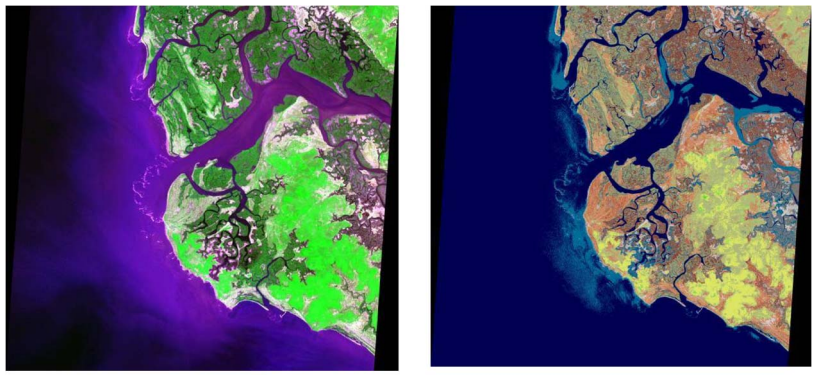
V. Ongoing and Future Developments

Part I of this paper, together with Sections II–IV, have proven that the proposed set of preliminary MS image classifiers, namely, LSRC, SSRC, AVSRC, AAVSRC, ISRC, and DSRC, identified as the integrated SRC system of classifiers, is operational, namely, it is effective, near real-time, fully automated



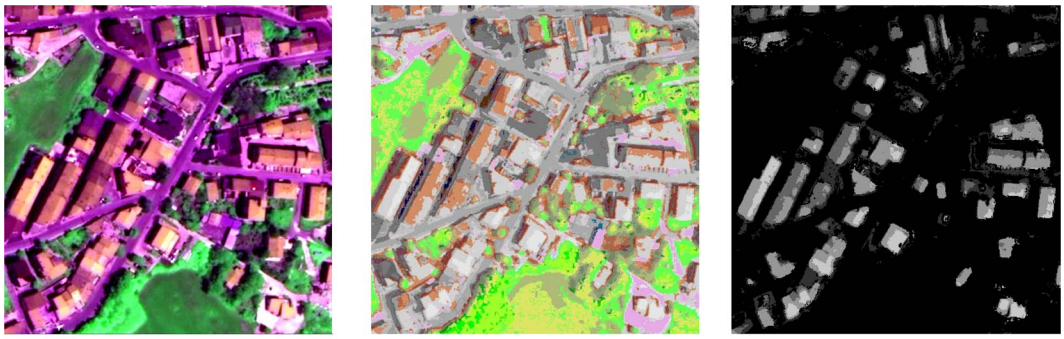
(a) (b)

Fig. 10. (a) AATSR image of the Black Sea (acquisition date: Oct. 10, 2006) in TOARF values, depicted in false colors (R: band CH4, G: band CH3, B: band CH1). (b) Output map, depicted in pseudocolors [the same as in Fig. 1(b)], generated by AASRC from the radiometrically calibrated image shown in (a).



(a) (b)

Fig. 11. (a) SPOT-2 high-resolution visible image in TOARF values of Senegal. False color image (R: band 3, G: band 2, B: band 1). Acquisition date: January 3, 2006, spatial resolution: 20 m. To be compared with Fig. 2(a) showing a SPOT-5 image of the same area of Senegal at a spatial resolution of 10 m. (b) Output map, depicted in pseudocolors [the same as in Fig. 1(b)], generated by DSRC from the radiometrically calibrated image shown in (a).



(a) (b) (c)

Fig. 12. (a) Zoomed image extracted from a QuickBird-2 image of the Campania region, Italy (acquisition date: June 13, 2004, 09:58 GMT), depicted in false colors (R: band CH3, G: band CH4, B: band CH1), 2.44-m resolution, calibrated into TOARF values and PAN-sharpened at 0.61-m resolution. (b) Output map, depicted in pseudocolors [the same as in Fig. 1(b)], generated by ISRC from the radiometrically calibrated image shown in (a). (c) Sunlit red roof mask generated from a stratified second-stage *ad hoc* spectral rule employing as inputs the image shown in (a) and the preliminary classification map shown in (b).

and robust to changes in the input data set acquired across time, space and sensors. These system characteristics make the integrated SRC system of systems eligible for use in a large variety of ongoing or future application fields as described as follows.

1) Development of a novel stratified relative calibration stage required by SPOT and Indian Remote Sensing Satellite (IRS) imagery to compensate for their lack of radiometric calibration offset parameters (see Section IV in Part I of this paper) in comparison with a radiometrically calibrated master MS image. For example, automated masking of water, crop fields, and clouds could be employed to search for invariant targets by the relative calibration approach proposed in [72].
2) Upscaling of the LSRC rule set to the superior spectral resolution of the MODIS satellite sensor, the ASTER satellite sensor, and aerial hyperspectral sensors such as the Specim AISA DUAL, the Itres SASI 600 and CASI 1500, and the Galileo Avionica SIM-GA.

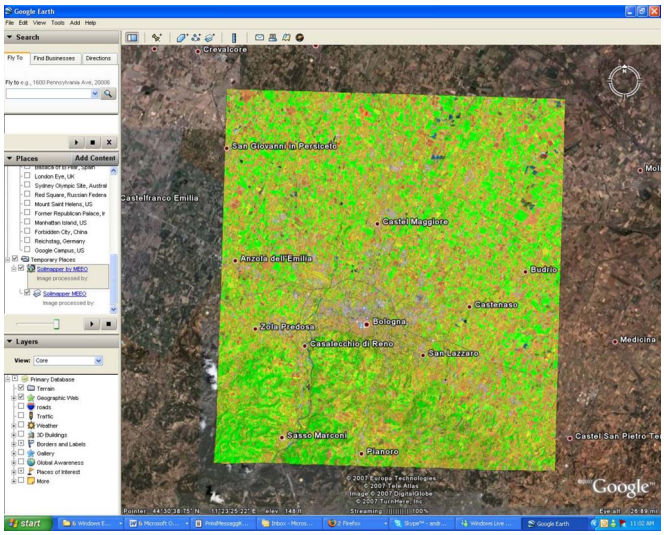


Fig. 13. Google Earth thematic layer consisting of a preliminary classification map which is as follows: 1) automatically generated from a radiometrically calibrated Landsat-7 ETM+ image (path 192, row 029, acquisition date: June 20, 2000) of Emilia-Romagna, Italy, by the integrated SRC system of classifiers and 2) transformed into the kml file format.

3) Augment the ISPRC rule set over a large number of VHR test images to take advantage of their fine spatial resolution which enables to see more objects, e.g., man-made objects such as roads, buildings, cars, etc. For example, Fig. 12(a) shows a zoomed image extracted from a QuickBird-2 MS image, radiometrically calibrated into TOARF values at 2.44-m resolution, which was PAN-sharpened at 0.61-m resolution with the ENVI software toolbox. Fig. 12(b) shows the preliminary classification map generated by ISRC from the PAN-sharpened image shown in Fig. 12(a). A red roof mask generated from a stratified second-stage class-specific spectral rule employing as inputs the MS image shown in Fig. 12(a) and the preliminary classification map shown in Fig. 12(b) is shown in Fig. 12(c).
4) Validation of the integrated SRC system of systems at continental and global scale in addition to local (areas up to 100 000 km^2) and regional (roughly between 100 000 and 1 000 000 km^2) geographic extents [67] as an alternative to the supervised (labeled) data-learning neural network classification approach adopted by the MODIS Science Team [15]. Global land cover product validation requires that a network of sites be developed. A global site network is an ambitious endeavor that requires cost sharing and interinstitutional cooperation [15].
5) Exploitation of the integrated SRC system of systems as a source of real-time public-domain geospatial thematic layers at either local, regional or global scale in commercial geobrowsers such as Google Earth, NASA's World Wind and the Microsoft Virtual Earth 3-D (see Fig. 13).
6) Use of the integrated SRC system in real-time applications such as:
   a) world wide web online customer service satisfaction on demand such as the ESA Service Support Environment portal that currently supports the generation of Landsat classification maps on demand [60];
   b) on-board integration of sensors, data processors, and communication systems in the so-called fourth-generation future intelligent EO satellites (FIEOS) [61], [62];
   c) developing operational satellite-based measurement systems [24, pp. 451–452] as envisaged by ongoing international programs such as GEOSS and GMES [2]–[4].
7) Exploitation of the integrated SRC system as an automatic source of thematic maps eligible for use in geospatial semantic querying of large-scale multisource RS image databases. This would represent a dramatic improvement over nonsemantic query modes currently available in image database retrieval systems based on text-based query strategies and query by either an image, object or multiobject example [63].
8) Development of a two-stage stratified hierarchical RS-IUS software toolbox employing SRC as its preliminary classification first stage. This RS-IUS implementation (refer to Fig. 3 in Part I of this paper) would become a potential alternative to existing subsymbolic pixel-based (e.g., ENVI [30]) or subsymbolic object-based (e.g., eCognition [64], refer to Fig. 2 in Part I of this paper) commercial RS image processing software toolboxes. In the computational framework of a two-stage stratified hierarchical RS-IUS architecture employing SRC as its preliminary classification first stage, several application examples are discussed below.
   a) A first-stage preliminary SSRC system in series with a second-stage stratified fuzzy classifier capable of combining color-based strata with context-based (either geometric, textural, or morphological) information performs an automatic separation of deciduous and evergreen broadleaf forest areas from crop fields and pastures in a 10-m resolution SPOT-5 image. First, a PAN (brightness) image is generated from the MS SPOT channels (CH) CH1 to CH4. Second, an SSRC spectral map is generated from bands CH1-CH4. Third, from the brightness (achromatic) image, multidisplacement multiscale texture contrast is computed (by means of a bank of second-order gray level cooccurrence matrices, GLCMs [65]) in cascade to multiscale morphological top hat of closing [66]. Fourth, crop field or pasture land covers are detected as a combination of: candidate vegetation spectral categories AND (Texture Contrast is Low) AND (Brightness is High). Finally, deciduous or evergreen broadleaf forests are detected as a combination of: candidate vegetation spectral categories AND [NOT(Crop field or Pasture)]. This procedure is shown in a zoomed area extracted from a SPOT-5 image of Belem, Brazil [see Fig. 14(a)–(f)].
   b) Exploitation of useful and computationally efficient second-stage topological filters when the target object's shape and size are known *a priori* and the target object is a bright object in a dark background, or vice versa. Valuable examples are the multiscale morphological top hat of either closing (see the previous

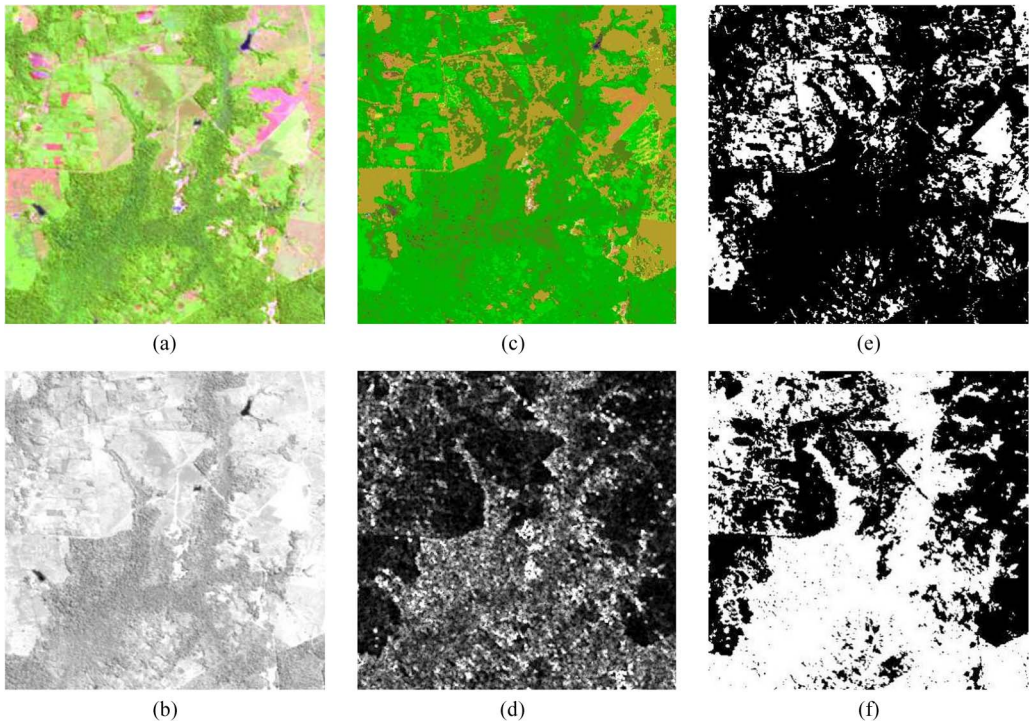


Fig. 14. (a) SPOT-5 image of Belem, Brazil, Path: 168, Row: 054, acquisition date: May 12, 2000. False color image (R: band 4, G: band 3, B: band 1). (b) PAN image generated from bands 1 to 4 of the SPOT-5 image. (c) SSRC spectral map generated from bands 1 to 4 of the SPOT-5 image and depicted in pseudocolors [the same as in Fig. 1(b)]. (d) Multidisplacement multiscale texture contrast in cascade to multiscale morphological top hat of closing extracted from the brightness image. (e) Crop field or pasture land covers, detected as a combination of: candidate vegetation spectral categories AND (texture is low) AND (brightness is high). (f) Deciduous or evergreen broadleaf forest land cover, detected as a combination of: candidate vegetation spectral categories AND [NOT(crop field or pasture)].

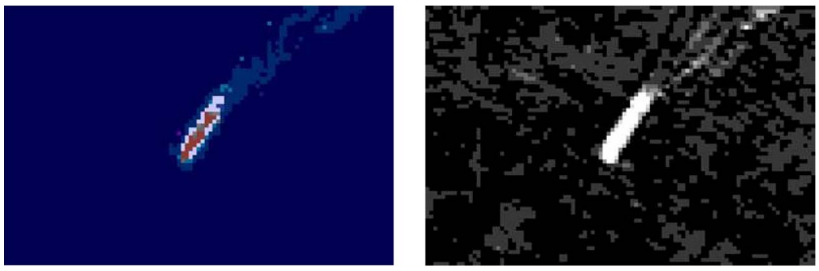
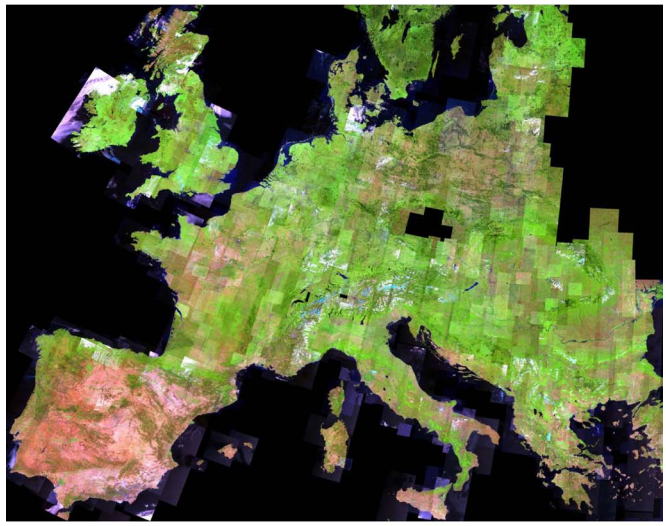


Fig. 15. (a) Preliminary spectral map, depicted in pseudocolors [the same as in Fig. 1(b)], generated from a SPOT-5 image of Belem, Brazil, Path: 168, Row: 054, acquisition date: May 12, 2000. Spatial resolution: 10 m. (b) Top hat of open, structuring element size: 5×5, generated from the brightness image extracted from a SPOT-5 image of Belem, Brazil, Path: 168, Row: 054, acquisition date: May 12, 2000. This morphological feature overlaps with the preliminary spectral classification map shown in (a).

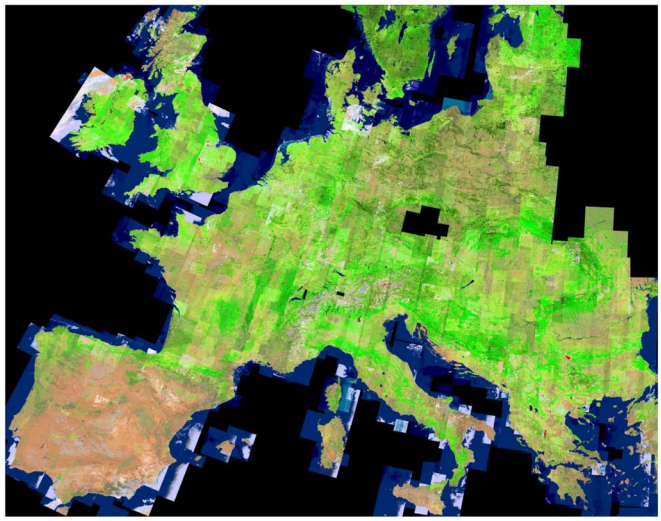
comment) or opening and the differential morphological profile [35]. For example, multiscale morphological top hat of opening is useful in detecting ships. In this case, the decision rule can be the following (also refer to Section II-C3b in Part I): a bright image object in a dark background, whose elongatedness is high, rectangularity is high, straightness of boundaries is high, whose length is between 3 and 300 m (e.g., equivalent to 3 to 300 pixels in a 1-m resolution sensor model) and whose spectral category is *either bare soil* or *built-up* and which is either surrounded by, adjacent to, or in the vicinity of water, is a ship (see Fig. 15).

c) Supplying LSRC with a real-time second-stage fire detection system suitable for MODIS imagery, as a potential alternative to the MODIS fire products where two spectral channels, centered on 4 and 11 μm, respectively, are employed [67]. The proposed real-time fire detection scheme would employ LSRC to generate a preliminary image partition into four strata: i) high vegetation biomass; ii) low vegetation biomass (e.g., herbaceous rangeland, shrubland); iii) bare soil or built up areas and iv) other classes inconsistent with fire (e.g., snow, clouds, water types). This stratification would be semantically richer than the water and cloud masking adopted by the MODIS fire products. Next, the proposed fire detection scheme would apply, on a stratified basis, the standard context-sensitive MODIS fire detection algorithm capable of separating candidate fire pixels from nonfire pixels. Finally, the Kaufman per pixel rule-based fire detection of: a) flaming fire; b) smoldering fire; and c) mixed fire would be applied, on a stratified basis, to candidate fire pixels [68].

d) Supplying AVSRC with a real-time second-stage fire detection system suitable for MSG imagery (geostationary, acquired every 15 min, where the two spectral channels, centered on 4 and 11 μm, respectively, employed by the MODIS fire products are available).
e) Providing SRC with a real-time second-stage flood detection system based on temporal changes in semantics driven from time sequences of preliminary classification maps.
f) Exploitation of the integrated SRC system to develop new second-stage methodologies for image mosaicking, e.g., to contribute to the GMES IMAGE2006 component and development of the European mosaic [69] (see Fig. 16).
g) Development of a semantic-based classification map compositing procedure starting from a multitemporal image sequence to recover from noisy data, e.g., due to cloud cover (see Fig. 17). To date, to compensate for atmospheric and aerosol scattering phenomena, classification of a multitemporal image sequence has generally combined multitemporal information at the signal level. For example, in [15], a time series of NDVI images is first composited by means of the max (Fuzzy-OR) operator. Then, the composited NDVI image is classified. This second multitemporal image classification strategy is justified exclusively by the fact that traditional single-date supervised classification of an image time series, which requires the collection of ground truth (reference) samples at each image acquisition time, is impractical.
h) Development of a two-stage RS image stratified topographic correction system capable of compensating for differences in solar illumination due to the irregular shape of the terrain. Some authors strongly discourage exploitation of current state-of-the-art RS image topographic correction methods where a single non-Lambertian surface reflectance model (in which reflectance depends on both observation and incident angles) is applied image-wide, although this model depends on surface roughness and is class-specific. In other words, the issue of RS image topographic correction is a typical chicken-and-egg dilemma: while image classification requires preliminary normalization of illumination conditions, the correction of topographic effects requires a classification map of the land surface as prior knowledge [70]. In this context, a preliminary spectral map can provide a multiple independent stratified (i.e., layer-specific) topographic correction system with kernel (reliable) image strata, [71] (see Fig. 18).
i) Development of a two-stage automatic stratified cloud detection system (refer to Fig. 19). It comprises, in cascade, a pixel-based preliminary classification first stage implemented as an integrated SRC system and a second-stage automatic and efficient prior knowledge-driven segment-based one-class classifier which incorporates the "stratified" or "layered" approach as follows. First, unlike ordinary inherently ill-posed segmentation in the raw image domain (refer to Section II-A in Part I of this paper), well-posed segmentation in the preliminary classification map domain is performed. As output, this well-posed segmentation module generates a segment description table where each segment is described in terms of [36]:
   i) a spectral-based semiconcept index (assigned by the first-stage SRC);
   ii) positional constraint equivalent to a minimum enclosing rectangle;
   iii) geometric attributes: area, perimeter, compactness, elongatedness, rectangularity, and straightness of boundaries;
   iv) spatial topological relationships between neighboring segments: adjacency and inclusion.

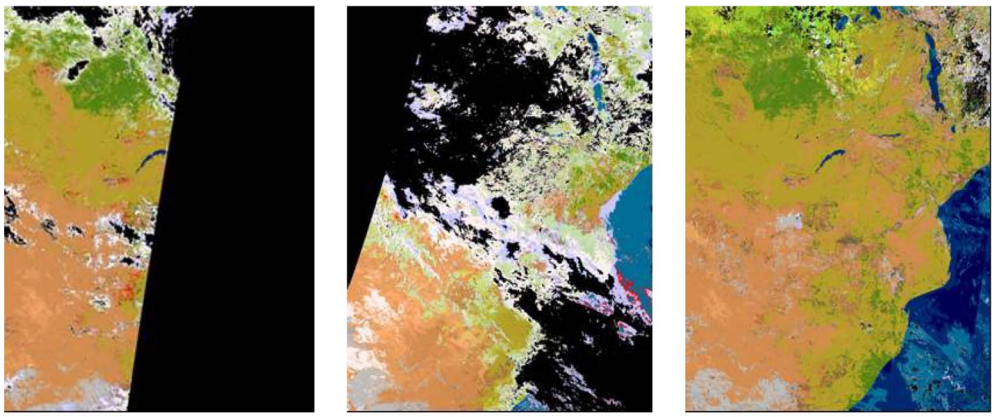


(a)



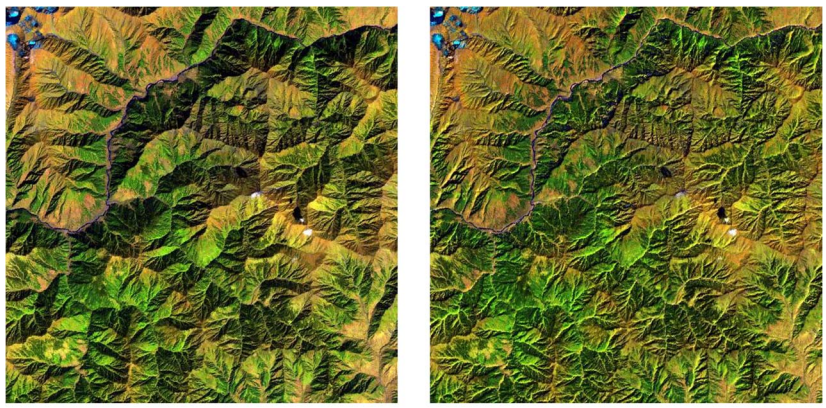
(b)

Fig. 16. (a) IMAGE2006 Coverage 1 mosaic, consisting of approximately 2000 IRS-P6 LISS-III, SPOT-4, and SPOT-5 images, mostly acquired during 2006, radiometrically calibrated into TOARF values and geometrically orthorectified. Images are depicted in false colors: red—band 4 (short-wave infrared), green—band 3 (NIR), blue—band 1 (visible green). Spatial resolution: 25 m. (b) Fully automated preliminary spectral classification of the IMAGE2006 Coverage 1 mosaic generated by SSRC and depicted in pseudocolors [the same as in Fig. 1(b)].



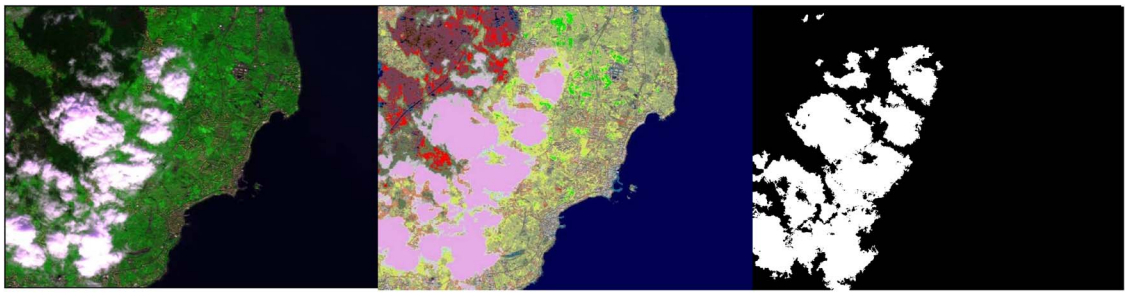
(a) (b) (c)

Fig. 17. (a) Preliminary spectral map, depicted in pseudocolors, generated from a SPOT-5 vegetation monitoring instrument (VMI) image of Zimbabwe acquired on December 1, 2006, spatial resolution: 1.1 km. (b) Preliminary spectral map, depicted in pseudocolors, generated from a SPOT-5 VMI image of Zimbabwe acquired on December 31, 2006, spatial resolution: 1.1 km. (c) Semantic-based composition, depicted in pseudocolors, of a sequence of single-date preliminary spectral maps generated from a multitemporal sequence of SPOT-5 VMI images of Zimbabwe acquired on December 2006, spatial resolution: 1.1 km. Courtesy of SARMAP s.a., urlinfo@sarmap.ch.



(a) (b)

Fig. 18. (a) Zoomed image extracted from a Landsat-7 ETM+ image of Pakistan, Path: 149, Row: 036, acquisition date: September 30, 2001. False color image (R: band TM5, G: band TM4, B: band TM1). (b) Stratified topographic correction of the image shown in (a) by means of a 16-class preliminary spectral classification map and a Shuttle Radar Topography Mission digital elevation model.



(a) (b) (c)

Fig. 19. (a) Zoomed image extracted from an advanced land observing satellite advanced visible and near infrared radiometer type 2 image of Sicily in false colors (R: band 2, G: band 4, B: band 1). (b) Output map, depicted in pseudocolors [the same as in Fig. 1(b)], generated by ISRC from the radiometrically calibrated image shown in (a). (c) Second-stage stratified cloud detector based on strata and segments detected in (b).

Second, a well-posed segment-based semantic-driven conditional region growing algorithm is applied in the classification map domain. Region growing criteria employ knowledge-driven constraints upon per segment color and geometric attributes and intersegment spatial topological relationships to detect clouds. For example, starting from a symbolic segment belonging to spectral category *core cloud* as a seed region, the region growing using semantics can grow onto adjacent symbolic segments belonging to spectral categories thin cloud over vegetation or thin cloud over water.

VI. Summary and Conclusion

One of the primary goals of the RS community traditionally involved with land cover and land cover change detection programs is the development of an operational flow of RS data and land cover information capable of providing advanced, reliable, and validated information products. The same objective, namely, the development of operational automatic RS-IUSs, is pursued by ongoing international EO programs such as the GEOSS and GMES. In these projects, the sustainability of integrated operational services based on EO data from multiple sources (satellite, airborne and *in situ*) [4] requires an international EO Cal/Val community-derived process to establish a quality assurance framework for harmonization and interoperability of EO data, metadata, derived information products and operations as outlined by the new QA4EO guidelines [2]. In this context, Part I of this paper presented five novel downscaled implementations of LSRC, an innovative operational fully automated spectral rule-based decision-tree classifier capable of mapping seven-band Landsat-like images (eventually synthesized from ASTER or MODIS spaceborne sensors) radiometrically calibrated into TOARF or surface reflectance values. LSRC together with its five downscaled implementations, namely, SSRC, AVSRC, AASRC, ISRC, and DSRC, is identified as the automatic integrated SRC system of systems requiring as input a radiometrically calibrated MS image acquired by almost any of the existing or future planned spaceborne optical imaging sensors in agreement with the QA4EO requirements.

In this paper, SSRC is tested both qualitatively and quantitatively in both synthetic and real-world data sets in comparison with LSRC, ISRC is compared against LSRC in three experiments exploiting one synthetic data set and two real-world data sets at local scale and, finally, AVSRC, AASRC, and DSRC are qualitatively (visually) assessed at continental and regional scale, respectively.

Conclusions of this experimental session are listed as follows.

1) In several real-world RS image classification problems (refer to Section II), a two-stage stratified RS-IUS implementation (refer to Section II-C3 in Part I of this paper) employing either LSRC or SSRC as its pixel-based preliminary classification first stage always achieves a classification accuracy superior to that gained by an alternative ordinary nonstratified single-stage classification approach. According to the theory of stratification, this result implies that LSRC and SSRC provide accurate image information symbolic primitives robust to changes in the input data set acquired across time, space, and sensors.
2) On a standalone basis, the integrated SRC system has been employed to enhance the consistency of a reference data set (acquired offline) with a given RS image (cleaning of reference samples per input image, refer to Section II-A9d).
3) In one synthetic test case (refer to Section II-B), due to the coarser spectral resolution of a SPOT-like image in comparison with a Landsat-like image irrespective of spatial resolution and intersensor differences in radiometric calibration, the SSRC loss in accuracy in a dichotomous vegetation/nonvegetation classification problem is estimated at about 1.5% with respect to LSRCs.
4) In one synthetic test case (refer to Section III-A), due to the inferior spectral resolution of an IKONOS-like image in comparison with a Landsat-like image irrespective of spatial resolution and intersensor differences in radiometric calibration, the ISRC loss in overall classification accuracy in comparison with LSRCs is estimated at about 10% (according to the confusion matrix kappa coefficient extracted from Table XV). The ISRC loss in accuracy of the dichotomous vegetation/nonvegetation classification problem is estimated at about 1% for spectral category *vegetation*, V, featuring a high LAI value, and about 100% for spectral category *weak rangeland,* WR, featuring a low LAI value (refer to Table XV), which is in line with theoretical expectations about spectral signature separability when a MIR channel is missing (refer to Section V in Part I). In other words, if spectral category WR is not considered, the ISRC separation of spectral category V from the rest of the world is 99% consistent with that generated from LSRC, which is a result far better than that theoretically expected (refer to Section V in Part I of this paper).
5) The qualitative assessment of the AVSRC, AASRC, and DSRC mapping results (refer to Section IV) is encouraging and in line with or even superior to theoretical expectations (refer to Section III in Part I of this paper). For example, DSRC is capable of mapping an input MS image consisting of only two bands, namely, one visible and a NIR channel.
6) Affected by the lack of radiometric calibration offset parameters, the absolute calibration of SPOT-5 imagery into TOARD and, as a consequence, TOARF values appears questionable in an experimental case (refer to Section II-A7), in agreement with Section IV in Part I. In practice, to make a SPOT-5 image well behaved, i.e., suitable for use in automatic classification systems, an additional relative calibration step (e.g., dark-object subtraction) is required in series with absolute radiometric calibration.

To summarize, Part I of this paper together with Sections II–V prove that the integrated SRC system of systems can be considered: 1) operational, namely, it is effective, near real-time, fully automated and robust to changes in the input data set acquired across time, space and sensors and 2) eligible for use as the automatic pixel-based preliminary classification first stage in a two-stage stratified hierarchical RS-IUS architecture. These conclusions are of potential interest to a large segment of the RS community involved with the development of operational satellite-based measurement systems as envisaged by the ongoing GEOSS and GMES international programs.

To conclude this paper, its authors wish to raise an open question about the degree of biological plausibility of the two-stage stratified hierarchical RS-IUS architecture employing the integrated SRC system of systems as its preliminary classification first stage. Despite its links to biological vision (mostly

omitted from Part I of this paper for the sake of simplicity; e.g., refer to Section II-C1 in Part I) and the scientific evidence collected to prove its effectiveness (refer to Sections II–V), the SRC color-based stratification mechanism is totally inapplicable to PAN imagery. On the contrary, the human visual system is perfectly capable of providing a symbolic description of the scene depicted in either color or PAN imagery. This evidence is more than enough to state that after this paper, the development of an effective, robust, and efficient artificial IUS remains an interdisciplinary, challenging and open problem.

ACKNOWLEDGMENT

The authors would like to thank the Editor-in-Chief, the Associate Editor, and anonymous reviewers for their helpful comments and their patience in reviewing this paper. The authors praise their competence and willingness to help.

REFERENCES

[1] A. Baraldi, V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino, "Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM + images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2563–2586, Sep. 2006.

[2] GEO, GEO 2007–2009 Work Plan: Toward Convergence, 2008. [Online]. Available: http://earthobservations.org

[3] GEO, The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan, adopted Feb. 16, 2005. [Online]. Available: http://www.earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf

[4] [Online]. Available: http://www.gmes.info

[5] L. Prechelt, "A quantitative study of experimental evaluations of neural network learning algorithms: Current research practice," *Neural Netw.*, vol. 9, no. 3, pp. 457–462, Apr. 1996.

[6] 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.

[7] [Online]. Available: http://www.ceos.org

[8] GEO/CEOSS, A Quality Assurance Framework for Earth Observation, Sep. 2008, ver. 2.0. [Online]. Available: http://calvalportal.ceos.org/CalValPortal/showQA4EO.do?section=qa4eoIntro

[9] [Online]. Available: http://www.coventry.ac.uk/ec/~nhunt/meths/strati.html

[10] M. L. Wilkie and S. Fortuna, *Status and Trends in Mangrove Area Extent Worldwide*. Rome, Italy: FAO—Forest Resour. Division, 2003.

[11] M. C. Cormier-Salem, *Les rivières du Sud. Sociétés et Mangroves Ouest-Africaines*, vol. 1/2. Paris, France: ORSTOM, 1999.

[12] F. Bertrand, "Mangrove dynamics in the Rivieres du Sud area—West Africa: An ecogeographic approach," *Hydrobiologia*, vol. 413, no. 1, pp. 115–126, Oct. 1999.

[13] E. S. Diop, A. Soumare, N. Diallo, and A. Guisse, "Recent changes of the mangroves of the Saloum River Estuary, Senegal," *Mangroves Salt Marshes*, vol. 1, no. 3, pp. 163–172, Sep. 1997.

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

[15] A. Strahler, D. Muchoney, J. Borak, M. Friedl, S. Gopal, E. Lambin, and A. Moody, "MODIS Land Cover Product Algorithm Theoretical Basis Document (ATBD)," MODIS Science Team, ver. 5.0. [Online]. Available: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod12.pdf

[16] A. Belward, The IGBP-DIS Global 1-km Land Cover Data Set DISCover, 1996, IGBP-DIS Working Paper.

[17] P. H. Swain and S. M. Davis, *Remote Sensing: The Quantitative Approach*. New York: McGraw-Hill, 1978.

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

[19] P. Smits, S. Dellepiane, and R. Schowengerdt, "Quality assessment of image classification algorithms for land cover mapping: A review and proposal for a cost-based approach," *Int. J. Remote Sens.*, vol. 20, no. 8, pp. 1461–1486, May 1999.

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

[21] E. Lunetta and C. Elvidge, *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. London, U.K.: Taylor & Francis, 1999.

[22] Y. V. Venkatesh and S. Kumar Raja, "On the classification of multispectral satellite images using the multilayer perceptron," *Pattern Recognit.*, vol. 36, no. 9, pp. 2161–2175, Sep. 2003.

[23] T. Lillesand and R. Kiefer, *Remote Sensing and Image Interpretation,* 3rd ed. New York: Wiley, 1994.

[24] G. Gutman, A. C. Janetos, C. O. Justice, E. F. Moran, J. F. Mustard, R. R. Rindfuss, D. Skole, B. L. Turner, II, and M. A. Cochrane, Eds., *Land Change Science*. Dordrecht, The Netherlands: Kluwer, 2004.

[25] X. Dai and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 5, pp. 1566–1577, Sep. 1998.

[26] [Online]. Available: http://glcfapp.umiacs.umd.edu:8080/esdi/index.jsp

[27] Landsat 7 Science Data Users Handbook. [Online]. Available: http://landsathandbook.gsfc.nasa.gov/handbook.html

[28] [Online]. Available: http://eospso.gsfc.nasa.gov/eos_homepage/for_scientists/atbd/docs/ASTER/BTS_ATBD.pdf

[29] [Online]. Available: http://www.spotimage.fr/html/_167_224_584_.php

[30] ITT Industries Inc., ENVI 4.3 User Manual, Boulder, CO, 2006.

[31] [Online]. Available: http://www.gds.aster.ersdac.or.jp/gds_www2002/service_e/release_e/g.t.u.p9_e.html

[32] V. J.-D. Tsai and Y.-T. Huang, "Automated image mosaicking," *J. Chin. Inst. Eng.*, vol. 28, no. 2, pp. 329–340, 2005.

[33] A. Svab and K. Kristof Ostir, "High-resolution image fusion: Methods to preserve spectral and spatial resolution," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 565–572, May 2006.

[34] A. Baraldi, L. Bruzzone, and P. Blonda, "A multi-scale expectation-maximization semisupervised classifier suitable for badly-posed image classification," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2208–2225, Aug. 2006.

[35] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[36] M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*. New York: Plenum, 1980.

[37] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon, 1995.

[38] V. Cherkassky and F. Mulier, *Learning From Data: Concepts, Theory, and Methods*. New York: Wiley, 1998.

[39] T. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.

[40] I. T. Nabney, *NETLAB: Algorithms for Pattern Recognition*. London, U.K.: Springer-Verlag, 2002.

[41] A. Baraldi, L. Bruzzone, P. Blonda, and L. Carlin, "Badly-posed classification of remotely sensed images—An experimental comparison of existing data labeling systems," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 1, pp. 214–235, Jan. 2006.

[42] [Online]. Available: http://www.ncrg.aston.ac.uk/netlab

[43] A. K. Shackelford and C. H. Davis, "A self-supervised approach for fully automated urban land cover classification of high resolution satellite imagery," in *Proc. 3rd Int. Symp. Remote Sens. Data Fusion Over Urban Areas—Urban*, Tempe, AZ, Mar. 14–16, 2005.

[44] A. K. Shackelford and C. H. Davis, "A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1920–1932, Sep. 2003.

[45] A. K. Shackelford and C. H. Davis, "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2354–2363, Oct. 2003.

[46] A. K. Shackelford and C. H. Davis, "Fully automated road network extraction from high-resolution satellite multispectral imagery," in *Proc. IGARSS*, Toulouse, France, Jul. 2003, vol. 1, pp. 461–463.

[47] Q. Jackson and D. Landgrebe, "An adaptive classifier design for high-dimensional data analysis with a limited training data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 12, pp. 2664–2679, Dec. 2001.

[48] J. T. Morgan, A. Henneguelle, J. Ham, J. Ghosh, and M. M. Crawford, "Adaptive feature spaces for land cover classification with limited ground truth," *Int. J. Pattern Recog. Artif. Intell.*, vol. 18, no. 5, pp. 777–800, Aug. 2004.

[49] K. W. Forsythe and P. Du, "Development patterns in Canada's largest urban agglomeration: Four decades of evolution," in *Proc. 1st EARSeL Workshop SIG Urban Remote Sens.*, Mar. 2–3, 2006. [Online]. Available: http://www2.hu-berlin.de/hu-geomatics/projects/earsel/CD_content/PDF/Session7_Forsythe.pdf

[50] M. P. Buchleim and T. M. Lillesand, "Semi-automated training field extraction and analysis for efficient digital image classification," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 9, pp. 1347–1355, Sep. 1989.
[51] G. Buttner, T. Hajos, and M. Korandi, "Improvements to the effectiveness of supervised training procedures," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 1005–1013, Jun. 1989.
[52] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. Int. Joint Conf. Artif. Intell.*, 1995, pp. 1137–1145.
[53] J. C. Bezdek, T. R. Reichherzer, G. S. Lim, and Y. Attikiouzel, "Multiple-prototype classifier design," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 28, no. 1, pp. 67–79, Feb. 1998.
[54] [Online]. Available: www.fabricadebani.ro/userfiles/GEO_press_release.doc
[55] A. Baraldi and K. Gutjahr, "Standardization of procedures for the quality assessment of panchromatic-sharpened multi-spectral imagery—Part I: Survey of the state-of-the-art," *IEEE Trans. Image Process.*, 2008, submitted for publication.
[56] M. Beauchemin and K. Thomson, "The evaluation of segmentation results and the overlapping area matrix," *Int. J. Remote Sens.*, vol. 18, no. 18, pp. 3895–3899, Dec. 1997.
[57] A. Di Gregorio and L. Jansen, *Land Cover Classification System (LCCS): Classification Concepts and User Manual*. Rome, Italy: FAO, FAO Corporate Document Repository, 2000. [Online]. Available: http://www.fao.org/DOCREP/003/X0596E/X0596e00.htm
[58] G. J. Hay and G. Castilla, "Object-based image analysis: Strengths, weaknesses, opportunities and threats (SWOT)," in *Proc. 1st Int. Conf. OBIA*, S. Lang, T. Blaschke, and E. Schöpfer, Eds., 2006. [Online]. Available: www.commission4.isprs.org/obia06/.../OBIA2006_Hay_Castilla.pdf
[59] B. Julesz, "Texton gradients: The texton theory revisited," *Biol. Cybern.*, vol. 54, no. 4/5, pp. 245–251, Aug. 1986.
[60] [Online]. Available: http://services.eoportal.org/portal/system/AboutUsUI.jsp
[61] G. Zhou and M. Kafatos, "Future intelligent earth observing satellites (FIEOS)," presented at the ISPRS/FIEOS Conf., 2002. [Online]. Available: http://www.isprs.org/commission1/proceedings02/paper/00031.pdf
[62] [Online]. Available: http://directory.eoportal.org/info_3rdInternationalSymposiumonFutureIntelligentEarthObservingSatellitesFIEOS.html
[63] C.-R. Shyu, M. Klaric, G. J. Scott, A. S. Barb, C. H. Davis, and K. Palaniappan, "GeoIRIS: Geospatial Information Retrieval and Indexing System—Content mining, semantics modeling, and complex queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 839–852, Apr. 2007.
[64] Definiens Imag. GmbH, eCognition User Guide 4, Munich, Germany, 2004.
[65] H. Anys and D.-C. He, "Evaluation of textural and multipolarization radar features for crop classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 5, pp. 1170–1181, Sep. 1995.
[66] P. Soille, *Morphological Image Analysis,* 2nd ed. Berlin, Germany: Springer-Verlag, 2003.
[67] C. Justice, L. Giglio, L. Boschetti, D. Roy, I. Csiszar, J. Morisette, and Y. Kaufman, "MODIS Fire Products Algorithm Technical Background Document," MODIS Science Team, Oct. 2006. ver. 2.3.
[68] Y. Kaufman, C. Justice, L. Flynn, J. Kendall, E. Prins, L. Giglio, D. Ward, W. Menzel, and A. Setzer, "Potential global fire monitoring from EOS-MODIS," *J. Geophys. Res.*, vol. 103, no. D24, pp. 32 215–32 238, 1998.
[69] GMES IMAGE2006 Mosaic. [Online]. Available: http://www.gmes-gseland.info/com/news/GMES-Land_NewsLetter_I1.00.pdf
[70] A. Baraldi, M. Gironda, and D. Simonetti, "Operational two-stage stratified topographic correction of spaceborne multispectral imagery employing an automatic spectral-rule-based decision-tree preliminary classifier," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 1, pp. 112–146, Jan. 2010.
[71] D. Riano, E. Chuvieco, J. Salas, and I. Aguado, "Assessment of different topographic corrections in Landsat-TM data for mapping vegetation types," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 5, pp. 1056–1061, May 2003.
[72] M. El Hajj, M. Rumeau, A. Begue, O. Hagolle, and G. Dedieu, "Radiometric normalization of high spatial resolution multi-temporal imagery: A comparison between a relative method and atmospheric correction," in *Proc. SPIE Eur.*, Florence, Italy, Sep. 17–20, 2007.
[73] O. Sjahputera, C. H. Davis, B. Claywell, N. J. Hudson, J. M. Keller, M. G. Vincent, Y. Li, M. Klaric, and C. R. Shyu, "GeoCDX: An automated change detection and exploitation system for high resolution satellite imagery," in *Proc. IGARSS*, Boston, MA, Jul. 6–11, 2008, pp. V-467–V-470, Paper FR4.101.1.

Andrea Baraldi was born in Modena, Italy, in 1963. He received the Laurea (M.S.) degree in electronic engineering from the University of Bologna, Bologna, Italy, in 1989. His master thesis focused on the development of segmentation and classification algorithms for remotely sensed optical imagery.

From 1989 to 1990, he worked as a Research Associate with the Centro di Studio per l'Interazione Operatore-Calcolatore, National Research Council (CNR), Bologna, and served in the army with the Istituto Geografico Militare, Florence, working on satellite image classifiers and geographic information systems (GIS). As a Consultant with the European Space Research Institute, European Space Agency, Frascati, Italy, he worked on object-oriented applications for GIS from 1991 to 1993. From December 1997 to June 1999, he was assigned with a postdoctoral fellowship in artificial intelligence with the International Computer Science Institute, Berkeley, CA. From 2000 to 2002, as a Postdoctoral Researcher, he joined the Global Vegetation Monitoring Unit, Institute for Environmental and Sustainability (IES), Joint Research Centre (JRC), European Commission, Ispra, Italy, where he worked on the development and validation of algorithms for forest classification of radar mosaics at continental scale. From 2005 to 2009, at the Spatial Data Infrastructure Unit, IES, he was involved with satellite optical image calibration, classification, and mosaicking at continental scale. Since his master thesis, he has continued his collaboration with Istituto di Scienze dell'Atmosfera e del Clima, Bologna, and Instituto di Studi sui Sistemi Intelligenti per I'Automazione, CNR, Bari, Italy. He is currently with Baraldi Consultancy in Remote Sensing, Bologna. His main interests center on image understanding, with special emphasis on the development of operational automatic hierarchical multisource multiresolution spaceborne image understanding systems consistent with biological vision.

Mr. Baraldi served as an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS from 2001 to 2006.

Laurent Durieux received the Ph.D. degree in geography, with a specialization in remote sensing and geographic information systems, from the University of Aix-en-Provence, Aix-en-Provence, France.

He spent two years with the Aerospace Technical Center, Sao José dos Campos, Brazil, to study the impact of deforestation on convective clouds. Later, for two years, he was with the Global Vegetation Monitoring Unit, Joint Research Centre, Ispra, Italy, to develop Earth observation techniques for global and regional vegetation mapping and monitoring with a particular emphasis on forest assessments. He is a Researcher with the Institut de Recherche pour le Développement, Montpellier, France. He is actually the Coordinator of the Monitoring of the Environment Assisted by Satellite (SEAS, in French) program in Brazil and organizes scientific cooperation with the Satellite Pour l'Observation de la Terre direct receiving antenna in Cayenne (French Guiana). His research interests focus on monitoring tropical land surface processes using data from Earth observing satellites.

Dr. Durieux is the Vice President of the French Representation of the South-American Association of Remote Sensing.

Dario Simonetti was born in Varese, Italy, in 1981. He received the Laurea (M.S.) degree in informatics from the Universita' dell'Insubria, Facolta' di Scienze Matematiche, Fisiche e Naturali, Varese, in 2004. His thesis focused on the development of improved classification algorithms for near-real-time fire detection at global scale from satellite optical imagery.

He is currently with the Global Environment Monitoring Unit, Institute for Environment and Sustainability, Joint Research Centre, European Commission, Ispra, Italy, where he was an Analyst–Programmer Consultant from 2004 to 2009. His main activities are focused on automatic near-real-time analysis of fire occurrences in African protected areas; development and assessment of processing chains for calibration, coregistration, topographic correction, and classification of satellite imagery; and development of user-friendly geographic information systems application for validation of satellite imagery classification across different epochs. He is contributing in delivering products and services derived from the analysis of satellite remote sensing data, with special emphasis on biodiversity and land cover change detection.

Giulia Conchedda was born in Nuoro, Italy. She received the Laurea (M.S.) degree in tropical agriculture from the University of Florence, Florence, Italy, where she specialized in remote sensing and geographic information systems applied to the monitoring of natural resources in tropical countries, and the Ph.D. degree in geography from the University of Louvain-la-Neuve, Louvain-la-Neuve, Belgium.

As a Consultant in spatial, environment, and food security analysis in Africa and Latin America countries, she has been working for several international organizations for more than ten years. She is currently with the European Commission, Brussels, Belgium.

Francesco Holecz received the M.Sc. degree in geography and the Ph.D. degree in remote sensing from the University of Zurich, Zurich, Switzerland, in 1985 and 1993, respectively.

From 1994 to 1996, he was a Visiting Scientist with the Radar Sciences and Engineering Section, NASA Jet Propulsion Laboratory, where his research focused on synthetic aperture radar (SAR) data calibration, polarimetry, and estimation of forest biomass using low-frequency polarimetric SAR data. From 1996 to 1999, he was a Research Scientist with the Remote Sensing Laboratories, Institute of Geography, University of Zurich, working on the development of an airborne X- and P-band interferometric SAR processor and spaceborne SAR applications in the domain of agriculture, forestry, and snow. As cofounder of SARMAP s.a., Purasca, Switzerland, he has been working as CEO of SARMAP s.a. since 1998. His main interests focus on the development of SAR and optical multispectral image-understanding algorithms for automatic mapping and inference of biophysical parameters.

Palma Blonda (M'93) received the Ph.D. degree in physics from the University of Bari, Bari, Italy, in 1980.

In 1984, she joined the Institute for Signal and Image Processing, Italian National Research Council, Bari, which has recently changed its name as the Instituto di Studi sui Sistemi Intelligenti per l'Automazione. Her research interests include digital image processing, fuzzy logic, neural networks, and soft computing applied to multisource remotely sensed data fusion and classification. She was recently involved with the project "Landslide Early Warning Integrated System (LEWIS)," EVG1-CT-2001-00055, founded by the European Community in the framework of Fifth PQ, where her research interests focused on multisource data fusion and classification techniques for the extraction of superficial changes of landslide-related factors observable from space.