



**HAL**  
open science

# Scale Object Selection (SOS) through a hierarchical segmentation by a multi-spectral per-pixel classification

Marco Chini, Alessandro Chiancone, Salvatore Stramondo

► **To cite this version:**

Marco Chini, Alessandro Chiancone, Salvatore Stramondo. Scale Object Selection (SOS) through a hierarchical segmentation by a multi-spectral per-pixel classification. *Pattern Recognition Letters*, 2014, 49, pp.214-223. 10.1016/j.patrec.2014.07.012 . hal-01065938

**HAL Id: hal-01065938**

**<https://hal.science/hal-01065938>**

Submitted on 18 Sep 2014

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

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

# **Scale Object Selection (SOS) through a Hierarchical Segmentation by a Multi-spectral Per-pixel Classification**

Marco Chini<sup>1,\*</sup>, Alessandro Chiancone<sup>2</sup>, Salvatore Stramondo<sup>3</sup>

<sup>1</sup>Centre de Recherche Public - Gabriel Lippmann, Belvaux, Luxembourg

<sup>2</sup>INRIA Grenoble Rhône-Alpes & GIPSA-Lab, Grenoble France

<sup>3</sup>Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy

## **\*Corresponding author:**

Marco Chini

Centre de Recherche Public - Gabriel Lippmann

Department of Environment and Agro-biotechnologies (EVA)

41, rue du Brill

L-4422 Belvaux, Luxembourg

Phone: +352 470261- 417

Fax : +352 470264

Email: [chini@lippmann.lu](mailto:chini@lippmann.lu)

## **Abstract**

In high resolution multispectral optical data, the spatial detail of the images are generally smaller than the dimensions of objects, and often the spectral signature of pixels is not directly representative of classes we are interested in. Thus, taking into account the relations between groups of pixels becomes increasingly important, making object-oriented approaches preferable. In this work several scales of detail within an image are considered through a hierarchical segmentation approach, while the spectral information content of each pixel is accounted for by a per-pixel classification. The selection of the most suitable spatial scale for each class is obtained by merging the hierarchical segmentation and the per-pixel classification through the Scale Object Selection (SOS) algorithm. The SOS algorithm starts processing data from the highest level of the hierarchical segmentation, which has the least amount of spatial detail, down to the last segmentation map. At each segmentation level, objects are assigned to a specific class whenever the percentage of pixels belonging to the latter, according to a pixel-based procedure, exceeds a predefined threshold, thereby automatically selecting the most appropriate spatial scale for the classification of each object. We apply our method to multispectral, panchromatic and pan-sharpened QuickBird images, considering two different test cases: a region on the Etna volcano (Italy), imaged at a 2.4 m resolution, and an area close to the town of Balakot (Pakistan), imaged at a 0.6 m resolution. On both test-cases the proposed approach enhanced the classification accuracy with respect to the single-segment per-pixel classification approach. A detailed analysis of the benefits achieved using the hierarchical segmentation with respect to a single segmentation is reported.

## **1. Introduction**

The classification of remote sensing images is a tool to help understanding and monitoring a large variety of scenarios, converting them into tangible information, which can be utilized in conjunction with other data sources. High Resolution (HR) images provide new opportunities in many environmental issues, raising at

the same time new problems in automatic classification procedures. Indeed, the aim of image classification analysis is to extract information and display the results in a simple and effective way for possible further studies. In HR images a single pixel represents a small area ( $0.5\div 10\text{ m}^2$ ) of the sensed scene and is usually smaller than the size of the objects of interest. Hence the information of a single pixel is strongly linked to the information carried by its neighbors, which are part of the same object, pointing out the necessity to perform an object-oriented analysis, although the measurable statistics typically exploited for classification purposes are necessarily pixel-based (Blaschke and Strobl, 2001; Blaschke, 2010). Several articles show different approaches to conjugate the necessity to exploit spectral and spatial information content. A possible approach to overcome the pixel as a reference unit for the analysis is referred to as the segmentation technique, which transforms an image by merging pixels that belong to “homogeneous” areas. The segmented image helps understanding the geometric structures of objects in the sensed scene, thus allowing any subsequent analysis to be based on segments made of several pixels rather than on single pixels. Segmentation techniques have been recently applied to build novel and efficient spectral-spatial classifiers considering segments rather than individual pixels (Tarabalka et al., 2009; Tarabalka et al., 2010a; Tarabalka et al., 2010b; Pulvirenti et al., 2011). A problem that is strongly correlated to segmentation is the scale of interest, as changing the scale changes the pattern of reality (Marceau and Hay, 1999). In HR images the presence of several possible scales makes classification a very complex procedure, making a hierarchical decomposition of the image necessary (Beaulieu and Goldberg, 1989). Hierarchical segmentation techniques process data by building a multi-level structure, in which objects are decomposed in sub-objects in the lower levels of the hierarchy. In other words, the multi-level structure presents a variety of possible scales of interest and adopting the correct scale of interest depending on the class is a challenging issue. Recently, papers in this field highlight the need to determine the optimal number of partitions of the data, since the hierarchy is usually too complex (Das et al., 2008; Gurrutxaga et al., 2010). Gorte and Gneletti, (2003) used a split and merge segmentation algorithm at different thresholds to produce a three-level segmentation hierarchy; in Akcay and Aksoy, (2008) connected components of

derivative morphological profiles (Pesaresi and Benediktsson, 2001) at different scales are investigated in order to produce a segmentation hierarchy, while the segments are labeled by using their spectral and textural content; binary partition tree (BPT) has also been used to segment hierarchically and classify hyperspectral and SAR images (Alonso-Gonzalez et al., 2013).

In this paper, the selection of the objects is carried out through a hierarchical segmentation by a supervised per-pixel classification, making use of optical QuickBird images of the Etna region (Italy) and of the village of Balakot (Pakistan). The purpose of this paper is to analyze the use of the entire hierarchical segmentation during classification, establishing whether or not improvements in classification accuracies are achieved analyzing different scales of interest, rather than considering only one segmentation map.

The rest of the paper is organized as follows. First we describe our methodology, which consists of merging unsupervised hierarchical segmentation with per-pixel supervised classification, namely Scale Object Selection (SOS); then we describe the obtained results applied to the Etna volcano and Balakot village test cases, highlighting the improvement with respect to per-pixel classification and the advantages of using the entire hierarchical tree with respect to a single scale of segmentation. Finally, we draw conclusions.

## **2. Methodology**

The method outlined for classifying HR optical images stems from the statement that human vision usually tends to generalize images into homogenous areas first, and then characterizes these areas more carefully (Gorte, 1998). In addition, Marceau and Hay (1999) stated that the scale is a window of perception and one should deal with several scales if the spatial resolution is finer than the size of the objects of interest, which is the case for many case studies of land cover classification using HR optical images. Indeed, HR multispectral remote sensing data have the potential for a more detailed and accurate mapping of the environment at meter ground resolution. At the same time, however, they present additional problems for information mining. In fact, due to their higher spatial resolutions, the relevant classes are composed of

various materials arranged in complex ways. At this spatial resolution, such materials do not usually have a direct relation with the spectral classes available. In this respect, since we are using multispectral images in the visible spectrum we should try to approximate the human eye perception in the best possible way (Marceau and Hay, 1999), providing land use maps that are suitable for our purpose. Thus, this problem is intrinsically related to the sensor resolution and cannot be solved by increasing the number of spectral channels (Pacifici et al., 2009). For these reasons, hierarchical approaches seem particularly suitable if we search for classes with different spatial scales in the data (Duda et al., 2001). Furthermore, this approach is convenient when the multispectral information content of each object is not so homogeneous, i.e. there is a contribution from a dominant spectral class but also from several others. On the other hand, supervised multispectral per-pixel classification approaches present some weaknesses when dealing with this kind of problems. Stemming from the above-stated considerations, we merge the results of a supervised multispectral per-pixel classification and an unsupervised hierarchical segmentation tree in order to classify objects at different scales. To this end we have implemented an algorithm composed of three main steps: 1) unsupervised hierarchical segmentation; 2) supervised per-pixel classification; and 3) final object classification, hereafter referred to as Scale Object Selection (SOS), in which for each object the most appropriate segments through the entire hierarchical tree are selected (step 1), making use of the supervised per-pixel classification map (step 2).

### 2.1. Hierarchical segmentation.

Usually, the main objective of image segmentation is to produce a set of non-overlapping image segments. The non-intersecting regions produced by the segmentation should be homogeneous, whereas the union of two adjacent regions should not be homogeneous (Pal & Pal, 1993). Among segmentation algorithms, the clustering technique (Jain, 2010), which is the multi-dimensional extension of the concept of thresholding, is the most immune to noise (Fu and Mui, 1981) and also represents a common way to segment imagery taking contextual information into account and dealing with multi-spectral data (Haralick and Dinstein,

1975; Pulvirenti et al., 2011). Hierarchical clustering techniques perform either a series of successive mergers or a series of successive divisions (Jain, 2010). Agglomerative methods start with individual objects, i.e. one pixel if no a priori information is given. This latter aspect could be a major limitation in terms of computation time if the algorithm were to be run on a per-pixel basis, since satellite images are usually composed of millions of pixels. An alternative way is to use an agglomerative hierarchical method starting from previously clustered data, so that our hierarchical segmentation is composed of the following steps:

i) k-means clustering algorithm (Jain, 2010); ii) Jeffries–Matusita (J–M) distance as an agglomerative rule (Richards and Jia, 1999); iii) segmentation of each clustered level (Pulvirenti et al., 2011).

Since we are more interested in the spatial arrangement of pixels rather than in their spectral information at pixel scale, we choose to work on Panchromatic images (PAN) or “pseudo” panchromatic images (Pseudo-PAN), i.e. the mean image of the four available multispectral bands.

Furthermore, in a segmentation procedure we are interested in separating homogeneous regions and structures with different spatial scale, therefore mathematical morphology can be profitably used (Soille, 2003), especially in a multiscale manner (Pesaresi and Benediktsson, 2001; Benediktsson et al., 2003; Benediktsson et al., 2005; Chini et al., 2009).

Starting from PAN/Pseudo-PAN images, we create the differential morphological profile (for further details on the morphological profile, please refer to Pesaresi and Benediktsson, 2001). Opening and closing morphological operators were applied to the PAN/Pseudo-PAN images, using square structuring elements with size ranging from  $3 \times 3$  up to  $49 \times 49$  pixels with a step of 2 pixels (i.e.,  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ , etc.). To reduce the input space, avoiding redundancy in the data yet retaining most of the variation, a principal component analysis is applied and the first effective Principal Components (PCs) are selected as the input space for hierarchical segmentation (Richards and Jia, 1999), which is composed by only the PCs that correspond to 99% of the variance ( $PC_{99\%}$ ).

Once the number of levels in which we want to decompose the image is fixed, we run the k-means algorithm, setting the number of clusters to the number of levels plus one (the highest level has 2 clusters). This value defines the depth of the tree. In this work we chose to use 50 clusters (49 levels). It is worth to point out that even if we over-segment the image, we are going to have a lower number of segments compared to a classical region-growing algorithm, which starts from the number of pixels that constitute the entire image. In order to build-up the entire hierarchy we calculate the J-M<sup>1</sup> distances for all cluster pairs, using the  $PC_{99\%}$ , in order to choose, for each iteration, the two clusters that are closest to each other. We then fuse such clusters and obtain a level with one cluster less. The procedure is iterated until a level with only two clusters is reached. It is important to highlight that in the agglomeration step, the contextual information of pixels and their geometrical patterns are taken into account, since the J-M distance is calculated by using the  $PC_{99\%}$  extracted from the morphological profiles, which account for the spatial arrangement of pixels.

Once the hierarchical clustering tree is obtained, each of the 49 levels between 2 and 50 are segmented (see Pulvirenti et al., 2011). At each level, the segmentation step transforms the clusters into segments that are composed by pixels having the following two properties: they belong to the same class (one of those labeled by the hierarchical agglomeration) and are contiguous in space. This is the step that actually takes the spatial contiguity of the pixels into account.

The pseudo-code of the hierarchical segmentation algorithm is summarized in the following:

---

<sup>1</sup> The J-M distance algorithm calculates the separability of two classes with respect to the input space, and it has output values  $x \in [0,2]$ , according to Richards and Jia, (1999). Values close to 2 indicate a higher degree of separability, while those close to 0 indicate a lower degree. Thus, if two clusters have a low degree of separability, it is likely that they have more similar characteristics. Hence, the J-M distance can be used to determine how clusters should be grouped to build the hierarchy.



*Input:*

$C_K$  - Clustering with  $K$  clusters

$PC$  - Principal Components of morphological profiles

*Output:*

$S=\{S_2, \dots, S_K\}$  - Hierarchical segmentation

$C=C_K$

for  $i=K$  to 2

$S_i=Seg(C, i)$  \*/ Segmentation of clustering  $C$  /\*

$J=J-M(C, i, PC)$  \*/ Matrix of JM distances between clusters/\*

$(a, b)=indmin(J)$  \*/  $(a, b)$  s.t.  $J_{a, b} = min(J)$  /\*

If  $a < b$  then \*/ Computing  $C_{i-1}$  clustering ./\*

$C(C=b)=a$  \*/assign  $a$  to each entry of  $C$  with value equal to  $b$ /\*

else

$C(C=a)=b$  \*/assign  $b$  to each entry of  $C$  with value equal to  $a$ /\*

end if

end for

---

## 2.2. Per-pixel classification

Maximum Likelihood (ML) classification is the most commonly used supervised classification method applied to remote sensing images (Richards and Jia, 1999). The ML classifier is parametric and relies on the second order statistics of a Gaussian probability density function model for the distribution of the feature vector of each class. It is often used as a reference for the comparison of classifiers, due to its optimality for Gaussian class probability density functions. In order to select training and validation samples,

both having statistical significance, and to avoid correlated neighbouring pixels, we adopted a Stratified Random Sampling (SRS) method, thus ensuring that even small classes, such as streets in this case, are adequately represented. This has been done randomly, by sampling the pixels in each class according to their areal extension, based on the available ground reference (Chini et al. 2008). For each class, 10% of pixels have been used to train the classifier (Training Set) and the remaining 90% have been used to assess the classification accuracy in terms of overall accuracy and K coefficient, (Validation Set) (Foody, 2002). The validation sets, both Etna and Balakot, have also been used to assess the results of all classifiers implemented in this paper.

### 2.3. Scale Objects Selection (SOS)

The selection of the scale of interest is a crucial point in order to correctly classify objects in the scene using HR optical images, which present more than one possible scale for the analysis. Indeed, an object can be more or less visible depending on the scale of interest. For this purpose a hierarchical representation of the image is very flexible and it can be applied to any type of image (Alonso-Gonzalez et al., 2013). The SOS algorithm explores the segmentation hierarchy by using a top-down strategy to automatically select the most representative level of segmentation. The combined use of a single segmentation map and a pixel-based classification has already been proved to enhance image classification (Tarabalka et al., 2009; Tarabalka et al., 2010b).

The input data of SOS are a hierarchical segmentation and a pixel-based classification. Given a Majority Voting Coefficient (MVC) value,  $0.5 < MVC < 1$ , SOS starts from the highest level of the hierarchical segmentation, which has the least amount of spatial detail, and proceeds down to the last segmentation map, classifying those segments with a percentage of pixels belonging to a specific class (assigned by the pixel-based classification) above the MVC. The tuning of the MVC parameter is performed during the training phase, selecting the one that gives the best results in terms of classification accuracy through cross-validation on the available training samples. MVC values between 0.5 and 1 have been tested with a

step of 0.05. Moreover, we want to underline that different MVC values lead to interesting classification maps that are not the best in terms of accuracy, but can enhance some characteristics of the image which are not considered in the selected ground truth (Fig. 2f). The output of SOS is a classification map in which segments are extracted through the whole segmentation hierarchy (see Fig. 1).

Let  $S$  be the segmentation hierarchy,  $P$  the pixel-based classification map with  $n$  classes and  $M$  the SOS output map. A segment of the  $i^{\text{th}}$ -level of the segmentation hierarchy is denoted by  $S_i^j \subset S$ . The segment  $S_i^j$  is split into one or more segments at the lower level of segmentation  $S_{i+1}$ ,  $\exists J = \{j_1, \dots, j_*\} \mid \bigcup_{j \in J} S_{i+1}^j = S_i^j$ . Let  $P^j$  and  $M^j$  be segments drawn in the pixel-based classification map and in the SOS output map.  $P^j$  and  $M^j$  enclose the portion of the sensed area defined by  $S_i^j$ . SOS considers a segment of the first level of the hierarchy  $S_0^j$  then creates a vector of the class frequency (e.g 30% water, 30% grass, 40% street) counting in the corresponding  $P^j$  the number of pixels belonging to the defined  $n$  classes. If the frequency of class  $f$  exceeds the MVC, SOS labels as  $f$  the segment  $M^j$ . If MVC is not exceeded by any class SOS proceeds to the lower level of  $S$  checking segments such that  $\bigcup_{j \in J} S_{i+1}^j = S_i^j$ . If SOS reaches the lowest level of  $S$ , the segment is labelled with the class  $f$  that maximizes the class frequency vector.

The pseudo-code for the SOS procedure is summarized in the following:

---

*Input:*

$S = \{S_2, \dots, S_K\}$  -Hierarchical segmentation

$P$ - Per pixel classification

MVC - Majority Voting Coefficient

*Output:*

$M$  - SOS classification

```

for i=2 to K
    for j=1 to |Si|
        VF=freq(Sij,P)           */ class frequencies in segment Sij /*
        [ma,f]= max(VF )         */ f.s.t ma=VFf /*
        if i=K OR ma>MVC then
            Mj=f
        end for
    end for
end for

```

---

SOS inserts spatial information in the classification at different scales of interest. This approach is also flexible, it can be used with different types of images or by adopting new pixel-wise classifications or different hierarchical segmentations approaches. Moreover, it is fast in building up the entire tree, since it fuses clusters at each iteration instead of fusing segments. The top-down strategy allows to rapidly classify large size classes, such as different kind of lava flows, which are present in the Etna test case or the River class in the Balakot one.

### 3. Results

In this section, we will discuss the results achieved with the SOS classifier for the two test cases selected, a region on the Etna volcano (Italy) and the Balakot village area (Pakistan), as well as the advantages of SOS compared to a per-pixel classifier in terms of classification accuracy. We will also highlight the advantages of a hierarchical segmentation approach compared to single segmentation, in terms of best spatial scale selection. Different spatial resolutions of the input data will be also tested; indeed for the Etna region we have a resolution of 2.4 m while for the Balakot village region is 0.6 m.

### 3.1 Etna (Italy)

Etna is an active volcano located in Eastern Sicily, Italy. Since the late 1970s, summit eruptions have shown an increase in their intensity. These events produce lava fountains that reach heights of several hundred meters with generation of abundant tephra and scoriae that can fall tens of kilometers away from the volcano. Etna's surface is highly dynamic, changing rapidly as lava flows are emplaced, ash falls deposit and pyroclastic cones grow. For these characteristics, the Etna volcano is periodically subject to remote sensing studies. In this work, the analysis has been carried out by using a HR QuickBird multispectral image acquired on May 27, 2006 and shown in Fig. 2a. The image has four bands with a geometric ground resolution of 2.4 m: three in the visible spectrum and one in the near infrared (blue, 0.45–0.52  $\mu\text{m}$ ; green, 0.52–0.60  $\mu\text{m}$ ; red, 0.63–0.69  $\mu\text{m}$ ; and near infrared, 0.76–0.90  $\mu\text{m}$ ). The four bands of the multispectral image constituted the input for the ML classifier. From these a Pseudo-PAN was formed, still at a 2.4 m geometric ground resolution, which was used as an input for the entire processing chain to obtain a hierarchical segmentation tree.

Eight different classes were identified, many of which are particular to this specific region: lava flows 2000-2005, lava flows 1800-1999, bright green vegetation, sparse vegetation, streets, cinder cones 1800-1999, dark green vegetation and cinder cones before 1800. The ground references for the scene, reported in Fig. 2b, have been obtained through a careful visual inspection of the multispectral QuickBird image and from the information provided by a geological map (Amici et al., 2011). The defined classes are reported in Table I, which contains the number of pixels for each class, as well as the colour table.

We started our processing by running the maximum likelihood classifier, thus obtaining the classification map shown in Fig. 2c. From a first visual inspection analysis, it is apparent how the land cover classes defined are not spectrally homogeneous, resulting in a fuzzy and noisy map. In order to quantitatively evaluate the results, the defined validation set has been used, and the overall accuracy and the K coefficient have been computed, obtaining 83% and 0.77 respectively. Subsequently, the aforementioned map was fused with the hierarchical segmentation in the SOS classification procedure, thus leading to the

classification map in Fig. 2d. This map has been obtained with an MVC of 0.8, which means that a segment is classified when the presence of a class from the ML exceeds 80%. It is worth noticing how such classification map is less noisy than the one obtained through the per-pixel ML classifier, with more homogeneous classified areas. Also in this case, the evaluation of the obtained map has been done quantitatively, allowing us to obtain an overall accuracy and a K coefficient of 90% and 0.87, respectively. Both coefficients have increased, thus highlighting the benefits of using the SOS classifier rather than a per-pixel one. The map obtained with an MVC of 0.8 is the one with the highest classification accuracy compared to the other MVC values tested. In Figs. 2e and 2f the maps obtained with MVC values of 0.95 and 0.6 are shown, respectively. These results highlight how lower MVC values privilege less noisy bigger classes, penalizing however the correct classification of smaller ones, while for higher values the performance tends to that of the per-pixel classifier. The MVC value, which provided the best classification map, represents a good compromise between noise reduction of the per-pixel classification, and the ability of classifying extended as well as small classes. Fig. 3a shows how many pixels have been selected in each level of the hierarchical segmentation. It is possible to notice how pixels have been selected in almost each level. This shows that, when dealing with land cover maps that have classes with different extension, we have to scan different scales of detail in order to select the best segmentation for each object.

In order to further stress the ability of the SOS approach to correctly classify classes of different extension, exploiting a hierarchical segmentation of the image, we have separately fused the ML map with all 49 segmented maps at different levels of detail. The results, in terms of accuracy, are shown in the graph in Fig. 3a. Accuracy is generally lower than the one obtained by using the SOS approach, except in one level, 34, where the accuracy slightly outperforms that of the SOS (K coefficient of 0.88). Looking at the general trend of the classification accuracy we can infer that it reaches a maximum value and then it starts to decrease, probably tending to the accuracy obtained with the classification at pixel scale.

If, on one hand, global coefficients for estimating the classification accuracy are good indicators for assessing the quality of a classification, on the other hand, in order to verify whether all classes have been

properly classified, it is better to look at the confusion matrix between ground truth and the classified map. We can compare the best single-segment classifier for this test-case, which, based on Fig. 3a, is obtained applying the ML classifier to segmentation level 34 (for which the highest overall accuracy is attained), to the best SOS result, obtained for an MVC of 0.8. The corresponding confusion matrices are provided in Fig. 4a and Fig. 4b respectively. It can be seen how the SOS classifies all classes with an accuracy of at least 77%, while the single-segment classifier reaches an accuracy of 70% in two classes and an even lower accuracy, namely 58%, for the Street class. The performance on the latter, for which an 88% accuracy is reached by exploiting the entire hierarchical scene decomposition, highlights the ability of SOS to classify indifferently dissimilar spatial scale classes in the image. Only two classes have a slightly higher accuracy with the single-segment approach, namely the most extended ones, thus causing the overall accuracy to be higher than that of the SOS. For the sake of clarity the classification map obtained with only one segmentation level is also shown in Fig. 5, where the above-mentioned considerations are highlighted, indeed the Street class has been almost under detected.

### 3.2 Balakot (Pakistan)

The second test case where the SOS algorithm was applied is located in a mountainous area, close to the Balakot (Pakistan) village (Chini et al., 2011). The area is composed by many land cover classes, some of which are also present in the other test case. The dataset consists of two QuickBird images taken on August 11, 2005, a PAN and a pan-sharpened (PS) one (Fig. 6a), both with a geometric resolution of 0.6 m. Unlike the Etna test case, where the definition of the classes was mostly based on existing geological maps, for this test-case the ground truth was entirely drawn by visual inspection of the PS image. Ten different classes were identified: rocks and sparse vegetation, grass, trees, streets, terraces, buildings, river, creek, roads and sand (see Table I and Fig. 6b).

The PS image was also used as input for the ML classifier, obtaining a classification map with a K coefficient of 0.86 and an overall accuracy of 88% (see Fig.6c). The PAN image instead was used as

input for the hierarchical segmentation procedure. Once the per-pixel classification map and the hierarchical segmentation structure were computed, they were in turn used as input of for the SOS procedure, testing different MVC values and selecting the one providing the best classification accuracy. In this case the best results were obtained with a 0.7 value, again providing an increase of classification accuracy respect to the ML per-pixel map (K coefficient of 0.90 and overall accuracy of 92%, see Fig. 6d).

To further highlight the benefit of using a multi-scale representation of the input images with respect to a fixed one, an additional experiment was carried out for this test-case, by providing a single segmentation level as an input to the SOS classifier. Therefore we obtained 49 classification maps running SOS on a single segmentation level. The results, in terms of accuracy, are shown in the graph in Fig. 3b. In this case the accuracy is always lower than the one obtained by using the SOS approach, while the general trend of the accuracy appears similar to that one of the Etna case: a maximum value is reached after which accuracy decreases towards figures obtained with a pixel based classification. Referring to the graphs in Fig. 3, we wish to stress that the accuracy obtained fusing only one segmentation level with the ML map, and the percentage of pixels classified in each level of the hierarchical tree, are directly linked. Indeed segmentation levels for which a considerable number of pixels have been selected by SOS correspond to significant variations in the classification accuracies of the single levels. This means that SOS takes advantage of all spatial scales that are strictly connected with those ones of the land cover classes under analysis.

#### **4. Conclusions**

Our Scale Object Selection (SOS) algorithm takes advantage of a hierarchical segmentation of the image, accounting for the geometrical characteristics of the objects within the scene, and of a per-pixel supervised classification, accounting for the multispectral information content. The hierarchical segmentation automatically provides a set of segmented maps of the image, with different levels of spatial detail. The final classification map is obtained by fusing segments at different spatial scales. The use of a majority voting coefficient in the segmentation hierarchy combined with the supervised per-pixel classification allows us to understand the spatial patterns that groups of pixels build up. The algorithm has been tested on two



different test cases, and in both cases this approach has enhanced the classification accuracy of the per-pixel maps. It is worth pointing out that if we use a single segmentation level the classification does not reach the highest accuracy, meaning that the method we propose allows to automatically select the most appropriate spatial scale of each class/object. The robustness of the algorithm has also been proved with respect to different spatial resolutions, namely 0.6 m and 2.4 m respectively for the two datasets of this study.

Finally, the power of this method for multisource information fusion should also be pointed out. Indeed, in this work we combined the hierarchical segmentation with a per-pixel classification extracted from the same dataset, i.e. Etna, or homogeneous type of data, i.e. Balakot. However, since the fusion is done in the post classification phase, there would be no obstacles to use a different data source, once the rules for the fusion have been defined.

## **Acknowledgment**

The authors would like to thank DigitalGlobe for providing the data used in this paper and are grateful to John Peter Merryman Boncori for proof-reading the manuscript. Marco Chini's contribution was supported by the National Research Fund of Luxembourg through the PAPAZZI project (CORE C11/SR/1277979).

## **References**

- Akçay, H.G., Aksoy, S., 2008. Automatic Detection of Geospatial Objects Using Multiple Hierarchical Segmentations. *IEEE Transactions on Geoscience and Remote Sensing* 46 (7), 2097-2111.
- Alonso-Gonzalez, A., Valero, S., Chanussot, J., Lopez-Martinez, C., Salembier, P., 2013. Processing Multidimensional SAR and Hyperspectral Images with Binary Partition Tree. *Proceedings of the IEEE* 101 (3), 723 – 747.

Amici, S., Piscini, A., Musacchio, M., Buongiorno, M.F., Neri, M., 2011. Geological Classification of Etna Volcano by Hyperspectral and Multispectral Satellite. Proc. Hyperspectral Imaging (HIS) conference 2011, 17-21.

Beaulieu, J., Goldberg, M., 1989. Hierarchy in Picture Segmentation: A Stepwise Optimization Approach. IEEE Transaction on Pattern Analysis and Machine Intelligence 11 (2), 150-163.

Benediktsson, J.A., Pesaresi, M., Arnason, K., 2003. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. IEEE Trans. Geosci. Remote Sens. 41 (9), 1940–1949.

Benediktsson, J.A., Palmason, J.A., Sveinsson, J., 2005. Classification of hyperspectral data from urban areas based on extended morphological profiles. IEEE Trans. Geosci. Remote Sens. 43 (3), 480–491.

Blaschke, T., Strobl, J., 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. GIS Zeitschrift fur Geoinformations systeme 14, 12-17.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 65, 2-16.

Chini, M., Pacifici, F., Emery, W.J., Pierdicca, N., Del Frate, F., 2008. Comparing statistical and neural network methods applied to very high resolution satellite images showing changes in man-made structures at Rocky Flats. IEEE Transactions on Geoscience and Remote Sensing 46 (6), 1812-1821.

Chini, M., Pierdicca, N., Emery, W.J., 2009. Exploiting SAR and VHR optical images to quantify damage caused by the 2003 Bam earthquake. IEEE Transactions on Geoscience and Remote Sensing 47 (1), 145 - 152.

Chini, M., Cinti, F. R., Stramondo, S., 2011. Co-seismic surface effects from very high resolution panchromatic images: the case of the 2005 Kashmir (Pakistan) earthquake. Natural Hazards and Earth System Sciences, 11, 931–943.

Das, S., Abraham, A., Konar, A., 2008. Automatic Clustering Using an Improved Differential Evolution Algorithm. *IEEE Transactions on Systems, Man, and Cybernetics — Part A: Systems And Humans* 38 (1), 218-237.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. *Pattern Classification*. 2nd ed. New York: Wiley.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment* 80, 185–201.

Fu, K.S., Mui, J. K., 1981. A survey on image segmentation. *Pattern Recognition* 13, 3–16.

Gorte, B. G. H., 1998. Probabilistic segmentation of remotely sensed images. PhD thesis, ITC Publication 63 (ITC, The Netherlands: Enschede).

Gorte, B.G.H., Gneletti, D., 2003. A method for object-oriented land cover classification combining Landsat TM data and aerial photographs. *Int. J. Remote Sensing*, 24 (6), 1273-1286.

Gurrutxaga, I., Albisua, I., Arbelaitz, O., Martin, J.I., Muguerza, J., Perez, J.M., Perona, I., 2010. SEP/COP: An efficient method to find the best partition in hierarchical clustering based on a new cluster validity index. *Pattern Recognition* 43, 3364–3373.

Haralick, R.M., Dinstein, I., 1975. A spatial clustering procedure for multi-image data. *IEEE Transactions on Circuits and Systems CAS-22*, 440–450.

Jain, A.K., 2010. Data Clustering: 50 Years Beyond K-Means. *Pattern Recognition Letters* 31 (8), 651-666.

Marceau, D.J., Hay, G.J., 1999. Remote sensing contributions to the scale issue. *Canadian Journal of Remote Sens.* 25 (4), 357-366.

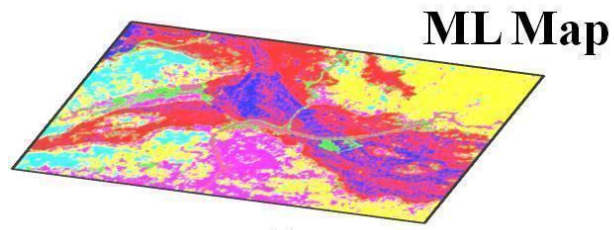
Pacifici, F., Chini, M., Emery, W.J., 2009. A neural network approach using multi-scale textural metrics from very high resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment*. 113 (6), 1276 – 1292.

- Pal, R., Pal, K. 1993. A review on image segmentation techniques. *Pattern Recognition* 26, 1277-1294.
- Pesaresi, M., Benediktsson, J.A., 2001. A new approach for the morphological segmentation of high resolution satellite imagery. *IEEE Trans. Geosci. Remote Sens.* 39 (2), 309–320.
- Pulvirenti, L., Chini, M., Pierdicca, N., Guerriero, L., Ferrazzoli, P., 2011. Flood monitoring using multi-temporal COSMO-SkyMed data: Image segmentation and signature interpretation. *Remote Sensing of Environment* 115 (4), 990-1002.
- Richards, J.A., Jia, X., 1999. *Remote Sensing Digital Image Analysis: An Introduction*. New York: Springer-Verlag.
- Soille, P., 2003. *Morphological Image Analysis, Principles and Applications*. 2nd ed. Berlin, Germany: Springer-Verlag.
- Tarabalka, Y., Benediktsson, J.A., Chanussot, J., 2009. Spectral–Spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques. *IEEE Transaction on Geoscience and Remote Sensing* 47 (8), 2973-2987.
- Tarabalka, Y., Chanussot, J., Benediktsson, J.A., 2010a. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition* 43 (7), 2367-2379.
- Tarabalka, Y., Benediktsson, J.A., Chanussot, J., Tilton, J.C., 2010b. Multiple spectral-spatial classification approach for hyperspectral data. *IEEE Trans. on Geoscience and Remote Sensing* 48 (11), 4122-4132.

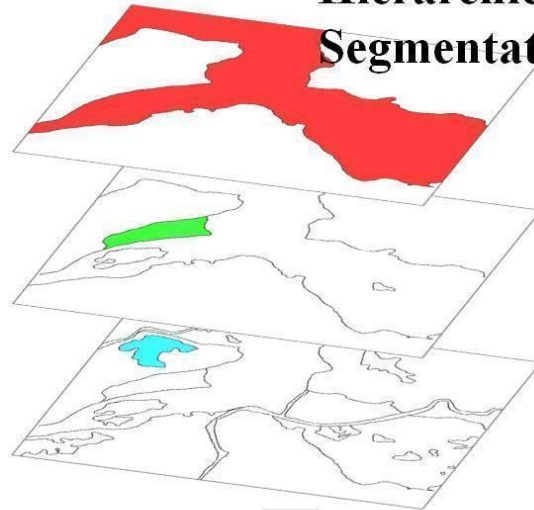
**Table I Etna and Balakot land cover classes.**

<b>Classes</b>	<b>Number of pixels</b>	
<b>Etna (Italy)</b>		
Lava flows 2000-2005	162.753	
Lava flows 1800-1999	43.661	

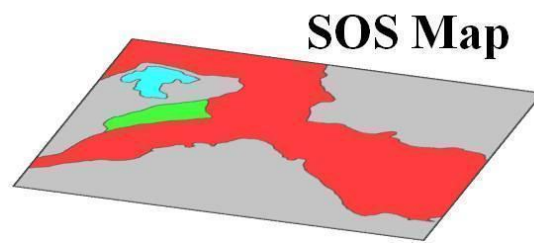
Bright green vegetation	301.284	Yellow
Sparse vegetation	146.567	Magenta
Streets	4.124	Red
Cinder cones 1800-1999	30.945	Blue
Dark green vegetation	34.338	Cyan
Cinders cones before 1800	5.388	Green
<b>Balakot (Pakistan)</b>		
Rocks and sparse vegetation	243.031	Red
Grass	73.937	Green
Trees	57.926	Blue
Streets	10.952	Yellow
Terraces	262.464	Cyan
Buildings	19.707	Red
River	154.005	Magenta
Creek	9.487	Green
Road	3.821	Purple
Sand	58.597	Orange

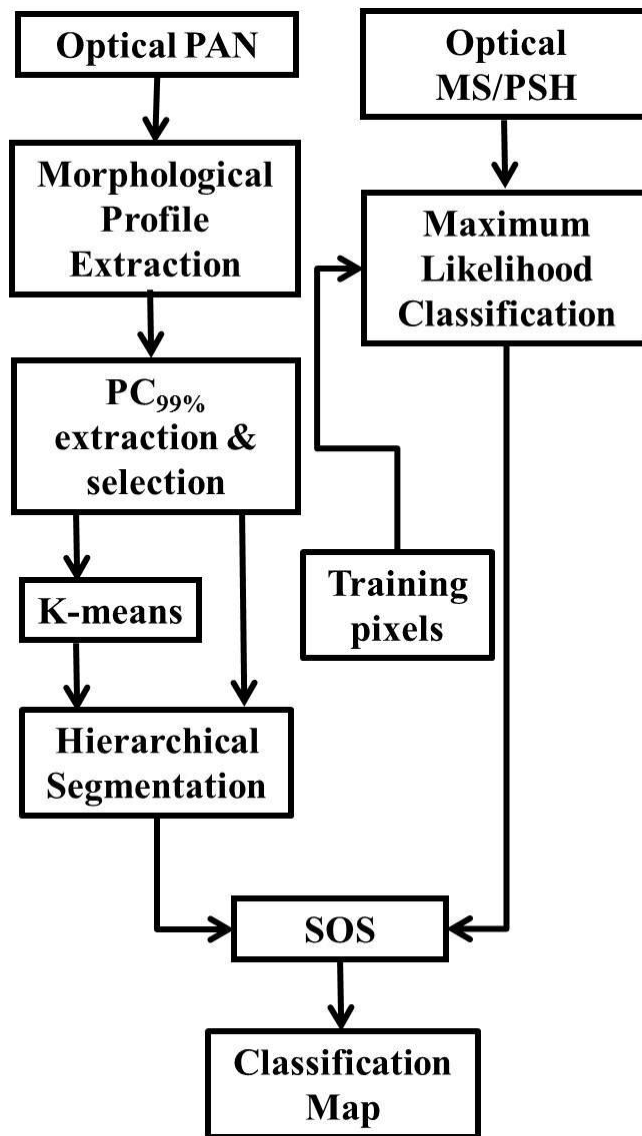


**+** Hierarchical  
Segmentation



**=**





a)

b)

Fig.1. a) Scale Object Selection (SOS) classifier scheme; b) Blocks diagram of the entire classification procedure.

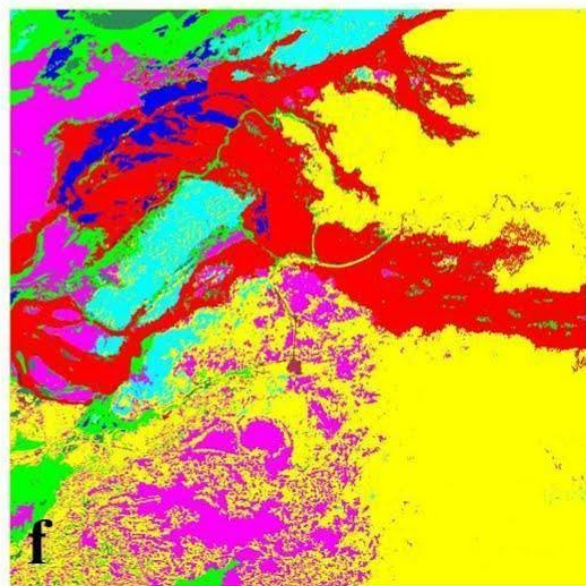
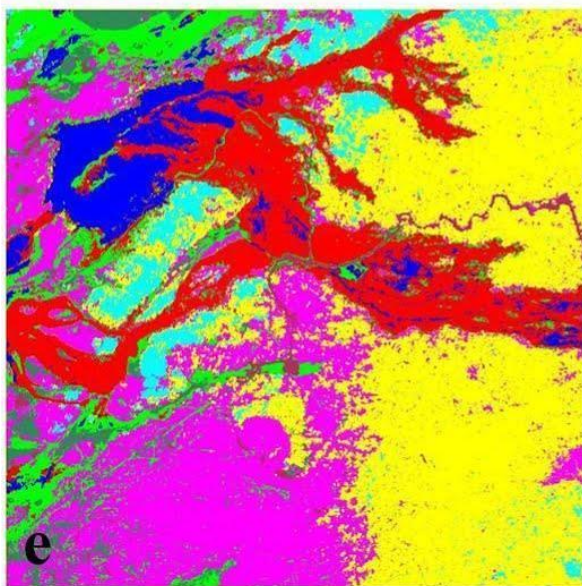
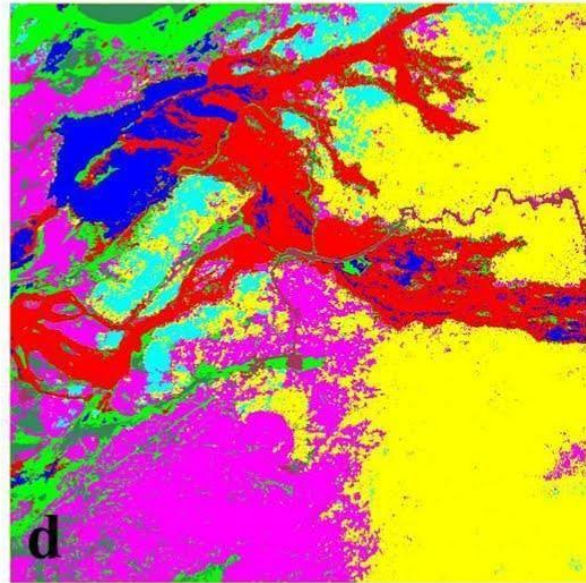
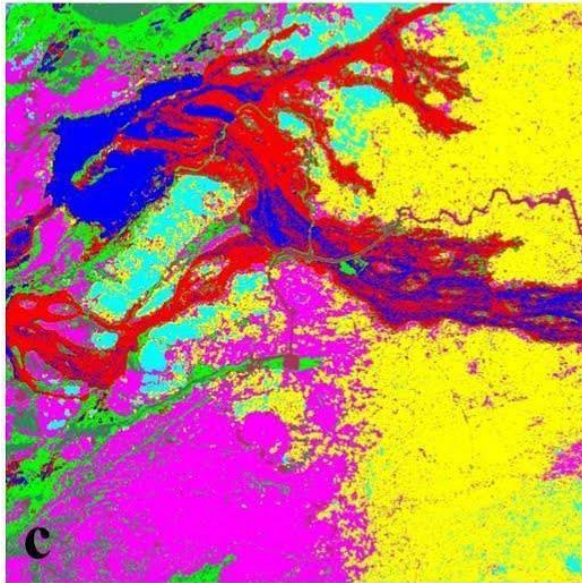
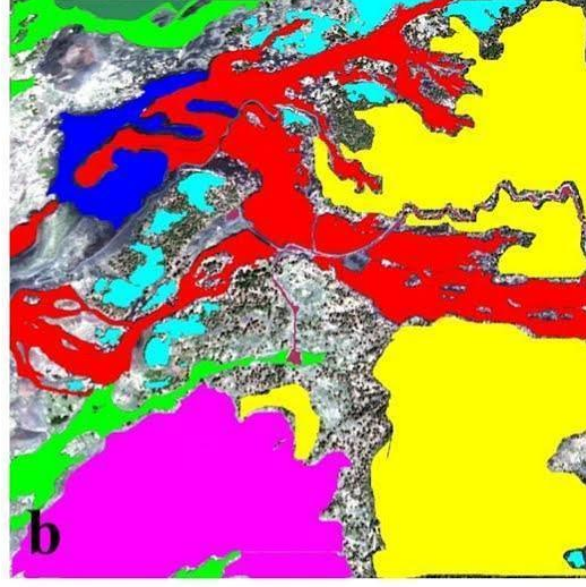
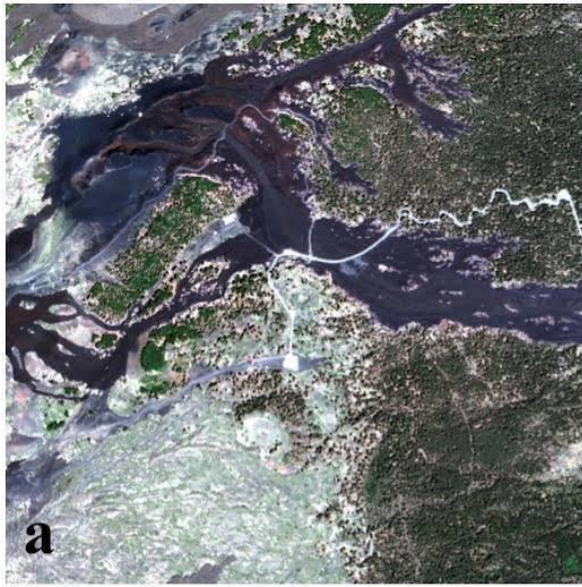
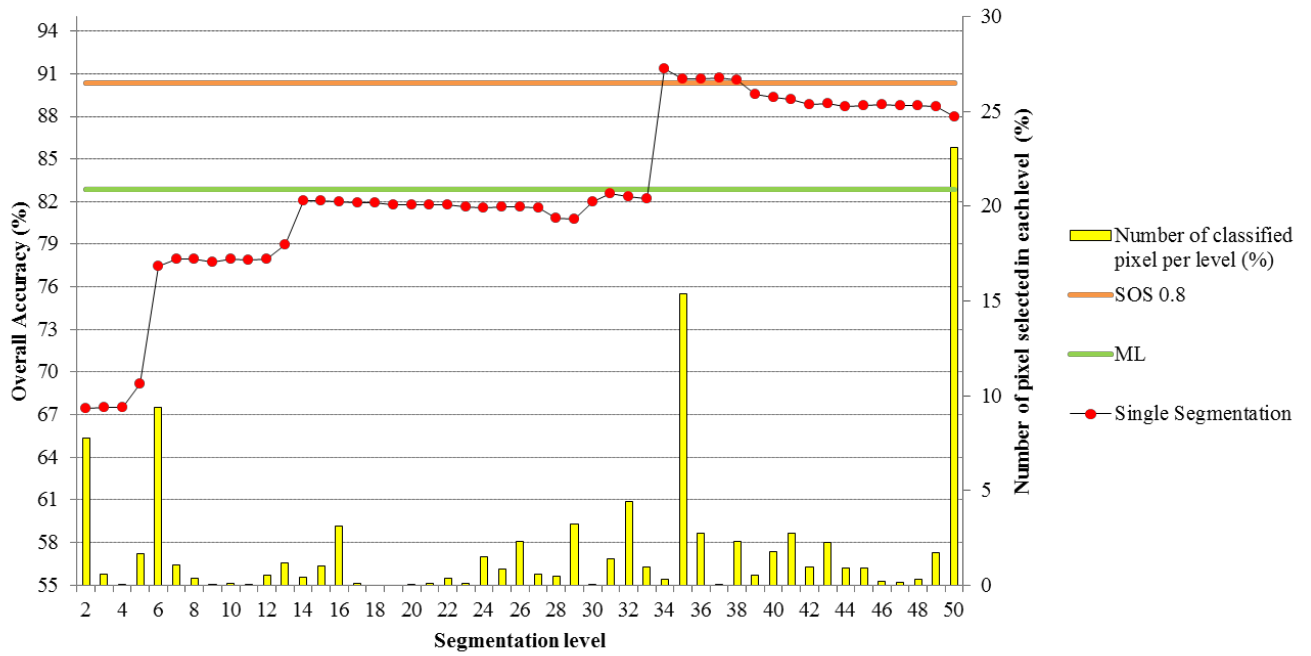




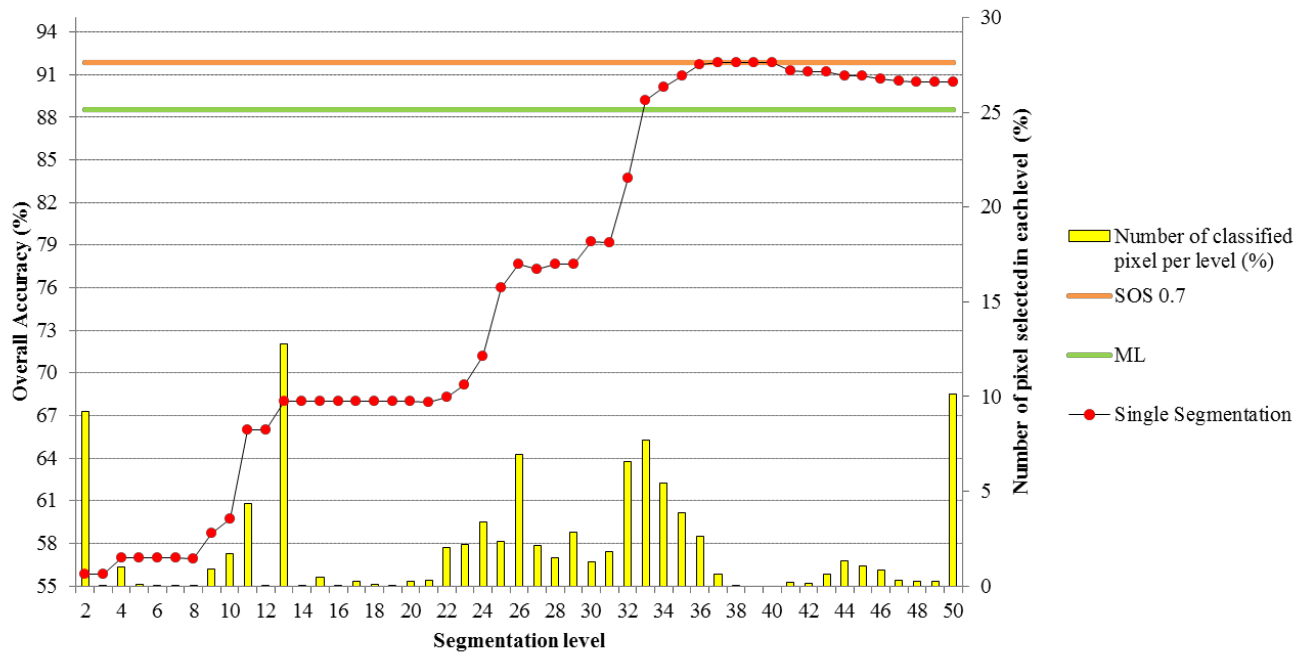
Fig. 2. a) Multispectral QuickBird image taken on May 27, 2006; b) ground reference map superimposed on the multispectral Quickbird image; c) per-pixel classification map obtained using the Maximum Likelihood classifier; d) SOS classification map, with MVC=0.8; e) SOS classification map, with MVC=0.95; f) SOS classification map, with MVC=0.6. The color legend refers to Table I.

### Overall Accuracy Etna



a)

### Overall Accuracy Pakistan



b)

Fig. 3. Classification accuracy obtained fusing each segmentation level with the per-pixel supervised classification (red dots) and the percentage of classified pixels in the different hierarchy levels for the two test cases are shown (yellow bars). Green line represents the classification accuracy obtained with the ML classifier, while the orange one is the accuracy obtained with SOS algorithm, where the MVC is 0.8 and 0.7 for the Etna and Pakistan test cases respectively. a) Etna test case; b) Pakistan test case.

Classes	1	2	3	4	5	6	7	8	Total
Lava flows 2000-2005, 1	125514 (85.69%)	1774	492	24	4	719	13	59	128599
Lava flows 1800-1999, 2	2990	2767 (70.44%)	533	2410	51	332	55	362	34412
Bright green vegetation, 3	743	248	266653 (98.34%)	942	1054	11	7624	121	277396
Sparse vegetation, 4	339	5409	2250	124147 (94.11%)	133	9	970	68	133325
Streets, 5	474	127	180	1506	2146 (57.81%)	49	3	1	4486
Cinder cones 1800-1999, 6	16052	1199	0	0	0	26692 (95.84%)	0	151	44094
Dark green vegetation, 7	169	54	933	87	312	0	22240 (71.96%)	0	23795
Cinder cones before 1800, 8	197	2805	115	2795	12	39	0	4088 (84.29%)	10051
<b>Total</b>	<b>146478</b>	<b>39295</b>	<b>271156</b>	<b>131911</b>	<b>3712</b>	<b>27851</b>	<b>30905</b>	<b>4850</b>	<b>656158</b>

a)

Classes	1	2	3	4	5	6	7	8	Total
Lava flows 2000-2005, 1	126090 (86.08%)	1720	473	24	4	614	13	57	128995
Lava flows 1800-1999, 2	2983	30355 (77.25%)	613	8912	72	330	83	368	43716
Bright green vegetation, 3	697	239	263140 (97.04%)	941	74	11	4920	119	270141
Sparse vegetation, 4	346	1158	4525	113458 (86.01%)	238	7	467	0	120199
Streets, 5	494	513	372	1175	3299 (88.87%)	51	47	0	5951
Cinder cones 1800-1999, 6	15534	1198	0	0	0	26798 (96.22%)	0	151	43681
Dark green vegetation, 7	102	40	1767	37	2	0	25375 (82.11%)	0	27323
Cinder cones before 1800, 8	232	4072	266	7364	23	40	0	4155 (85.67%)	16152
<b>Total</b>	<b>146478</b>	<b>39295</b>	<b>271156</b>	<b>131911</b>	<b>3712</b>	<b>27851</b>	<b>30905</b>	<b>4850</b>	<b>656158</b>

b)

Fig. 4. a) confusion matrix of the classification map obtained fusing the segmentation level 34 with the per-pixel supervised classification; b) confusion matrix of the SOS classification map with a MVC=0.8.

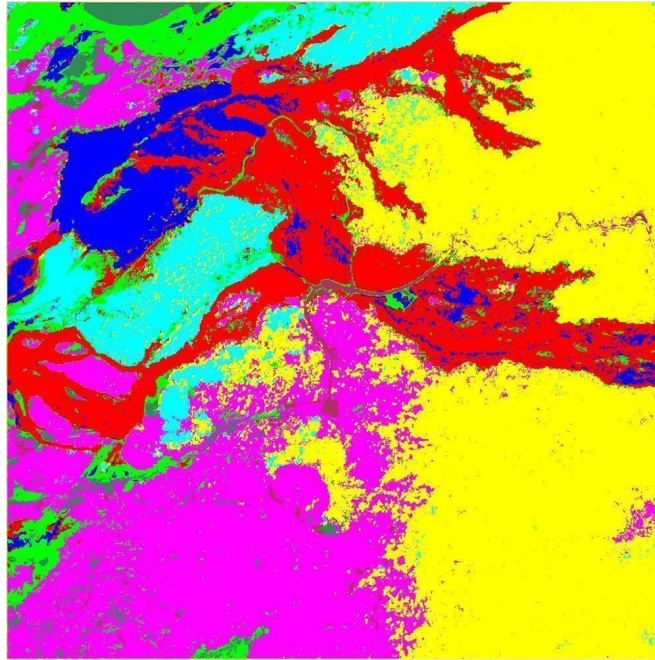


Fig. 5. Classification map obtained by fusing the segmentation level 34 with the per-pixel supervised classification.

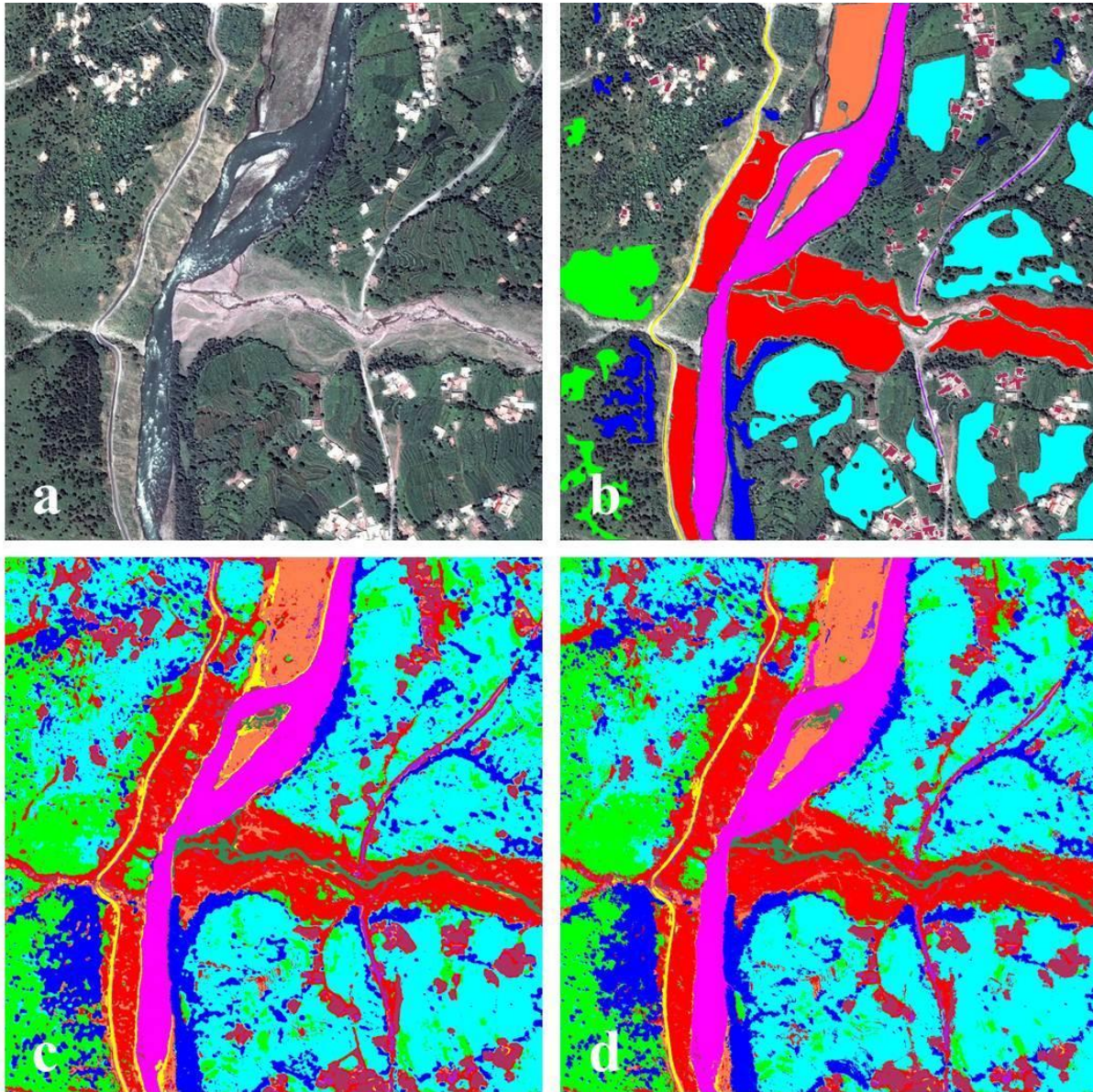


Fig. 6. a) Pan-sharpened QuickBird image near the Balakot (Pakistan) village taken on May 27, 2006; b) ground truth map superimposed on the pan-sharpened Quickbird image; c) per-pixel classification map obtained using the Maximum Likelihood classifier; d) SOS classification map, with MVC=0.7. The color legend refers to Table I.