*Article*

# Fast Segmentation and Classification of Very High Resolution Remote Sensing Data Using SLIC Superpixels

## Ovidiu Csillik

Department of Geoinformatics—Z_GIS, University of Salzburg, 5020 Salzburg, Austria;
ovidiu.csillik@sbg.ac.at; Tel.: +43-662-8044-7559

**Abstract:** Speed and accuracy are important factors when dealing with time-constraint events for disaster, risk, and crisis-management support. Object-based image analysis can be a time consuming task in extracting information from large images because most of the segmentation algorithms use the pixel-grid for the initial object representation. It would be more natural and efficient to work with perceptually meaningful entities that are derived from pixels using a low-level grouping process (superpixels). Firstly, we tested a new workflow for image segmentation of remote sensing data, starting the multiresolution segmentation (MRS, using ESP2 tool) from the superpixel level and aiming at reducing the amount of time needed to automatically partition relatively large datasets of very high resolution remote sensing data. Secondly, we examined whether a Random Forest classification based on an oversegmentation produced by a Simple Linear Iterative Clustering (SLIC) superpixel algorithm performs similarly with reference to a traditional object-based classification regarding accuracy. Tests were applied on QuickBird and WorldView-2 data with different extents, scene content complexities, and number of bands to assess how the computational time and classification accuracy are affected by these factors. The proposed segmentation approach is compared with the traditional one, starting the MRS from the pixel level, regarding geometric accuracy of the objects and the computational time. The computational time was reduced in all cases, the biggest improvement being from 5 h 35 min to 13 min, for a WorldView-2 scene with eight bands and an extent of 12.2 million pixels, while the geometric accuracy is kept similar or slightly better. SLIC superpixel-based classification had similar or better overall accuracy values when compared to MRS-based classification, but the results were obtained in a fast manner and avoiding the parameterization of the MRS. These two approaches have the potential to enhance the automation of big remote sensing data analysis and processing, especially when time is an important constraint.

**Keywords:** oversegmentation; runtime; OBIA; computer vision; pixels; superpixels; random forest

## 1. Introduction

In the last decades, object-based image analysis (OBIA) has emerged as a sub-discipline of GIScience devoted to analysis and processing of very high resolution (VHR) satellite imagery [1]. OBIA builds on older concepts of image analysis, like image segmentation and classification [2,3]. Image segmentation aims to partition relatively homogeneous image objects, non-overlapped and spatially adjacent [2]. A 'perfect' segmentation should aim to minimize internal variation of image objects and maximize external difference from neighbors. Many image segmentation algorithms have been developed [4], but segmentation is still an unresolved problem in OBIA [1,5,6]. This is due to the fact that segmentation is sensitive to many factors, like image sensor resolution, scene complexity or number of bands [6]. Object-based classification of VHR images represents a viable

alternative to the traditional pixel-based approach [7], minimizing the intra-class spectral variation using objects [8]. Numerous studies have addressed comparative analysis between pixel-based and object-based classification [8–13]. Besides many advantages [8], the main drawback of object-based classification is the dependency of the final accuracy on the quality of the segmentation results. Therefore, achieving a desirable degree of accuracy for segmentation and/or classification using objects require a large amount of time and parameterization of algorithms.

Many existing segmentation algorithms in OBIA, including the popular multiresolution segmentation (MRS) [14], use the pixel-grid as the underlying representation. Pixels are the atomic units of a raster, but they are not natural entities and, therefore, are unlikely to match the content of the space represented [15]. It would be more natural and efficient to work with perceptually meaningful entities that are derived from pixels using a low-level grouping process [16]. Therefore, we can partition an image into superpixels, which are the result of perceptual grouping of pixels based on similar characteristics (e.g., color) [16]. The term superpixels was first introduced by Ren and Malik [17], when they used them as a preprocessing stage to organize an image into superpixels through oversegmentation, thus simplifying the computation in later stages. Using superpixels instead of pixels in the segmentation process has certain advantages: (1) superpixels carry more information than pixels and adhere better to the natural image boundaries [16,18]; (2) superpixels have the scale between the pixel level and the object level [19,20]; (3) they are of low computational complexity and can speed-up the subsequent image processing [17–19] and (4) superpixels reduce the susceptibility to noise and outliers, capturing image redundancy [21].

In computer vision, using superpixels or perceptually meaningful atomic regions [19] to speed up later-stage visual processing is becoming increasingly popular in many applications [16,19,22]. Achanta, et al. [19] broadly categorized the algorithms of generating superpixels as either graph-based or gradient ascent methods and compared five state-of-the-art methods, in terms of speed, ability to adhere to image boundaries, and impact on segmentation performance. Graph-based algorithms generate superpixels by minimizing a cost function defined over the graph. The most known algorithms of this category are the normalized cuts algorithm [23], efficient graph-based image segmentation [24], regular superpixel lattice [25], and energy optimization framework [26]. Gradient-ascent-based algorithms start from a rough initial clustering of pixels and iteratively refine the clusters until some criterions are met to form the superpixels [19]. This includes mean-shift algorithm [27], quick-shift [28], watershed approach [29], and turbopixel method [30]. Achanta, et al. [19] proposed a new superpixel algorithm: Simple Linear Iterative Clustering (SLIC), which is an adaptation of *k*-means clustering for superpixel generation, but faster, more memory efficient, adheres well to the boundaries and improves the performance of segmentation algorithms.

In remote sensing domain, superpixels are used in a diverse range of applications and data types, like optical imagery [31,32], SAR images [33] or hyperspectral images [34,35]. Guangyun, et al. [20] used watershed segmentation to produce superpixels and further used them in a superpixel-based graphical model for remote sensing image mapping. Ortiz Toro, et al. [36] used superpixels to define the neighborhood similarity of a pixel adapted to the local characteristics of each image, in order to improve an unsupervised segmentation algorithm for multispectral imagery. Vargas, et al. [37] developed a new scheme for contextual superpixel description based on the bag-of-visual words. Garcia-Pedrero, et al. [38] combined edge-based and superpixel processing for automatic generation of thematic maps for small agricultural parcels. Fourie and Schoepfer [31] introduced a sample supervised segment generation method which made use of low-level image processing functions (SLIC superpixels). Stefanski, et al. [39] introduced the Superpixel Contour algorithm in remote sensing applications, in order to optimize a semi-automatic object-based classification of multitemporal data using Random Forest.

Nowadays, an increasing number of satellites deliver a huge amount of remote sensing data that needs to be processed and analyzed, driven by specific purposes. Spatial science is playing an important role in tackling the big problems of humanity, like climate change, sustainable development,

mobility, health, as well as safety and security and for this we need to deliver high-quality spatial information (i.e., timely, complete, reliable, geometrically and thematically accurate) [40]. Speed and accuracy are important factors when dealing with big remote sensing data or time conditioned events (e.g., earthquakes, tsunamis, landslides, refugees, tornados, oil spills etc.) for damage, disaster, risk and crisis-management support [2,41–43]. If a satellite image is increasing in landscape complexity, number of bands and extent, a high amount of time and computational resources are needed in order to automatically extract or classify features of interest. To our knowledge, no study has tested the idea of generating SLIC superpixels a priori of MRS to improve the computational time or to assess the performance of SLIC superpixels for object-based classification of remote sensing data.

In this study, we followed two research designs: (1) usage of SLIC superpixels in the process of segmenting VHR satellite images and (2) comparing SLIC superpixel classification with traditional pixel-based and object-based classifications. In the first scenario, we tested a workflow for image segmentation of VHR images, starting the MRS from the superpixel level and aiming at reducing the amount of time and computational resources needed to partition a relatively large datasets of VHR remote sensing data. This approach is compared with the traditional one, starting the MRS from the pixel level, regarding the geometric accuracy of the objects and the computational time. In the second scenario, we compared the classification accuracy of different SLIC superpixel sizes with pixel-based and object-based classifications. The main goal was to evaluate whether an oversegmentation obtained through SLIC superpixels is sufficient enough for a Random Forest classifier in order to obtain a classification accuracy similar to conventional object-based classification.

The next section (Section 2) describes the datasets, the theoretical background of the SLIC superpixels, generation of objects, random forest classifier, and the accuracy measures for segmentation and classification. Section 3 compares the results in terms of accuracy and computational time. Section 4 drafts the major implications of the results, analyzes the performance and limitations of the study and states further developments. Section 5 concludes the main findings of our study.

## 2. Materials and Methods

### 2.1. Datasets

Tests were conducted on very high resolution remote sensing data, differing in spatial resolution, number of bands, extent, and landscape (Table 1, Figure 1). Test area T1 covers a dense residential and services area, being located in the area of the main train station in Salzburg, Austria. The T2 test area covers the southern part of the city of Salzburg, being comprised of dense clustered residential and commercial buildings with large green spaces in between. Test area T3 represents a sensitive riparian habitat with forests, agricultural fields, and water bodies, situated at the border between Austria and Germany, in the northern part of the city of Salzburg [44].

The spatial resolution ranges from 0.5 m (T3) to 0.6 m (T1 and T2), respectively, as well as the number of bands, between four (T1 and T2) and eight (T3) (Table 1). We have used test areas with different characteristics to assess how the computational time for segmentation is influenced by these factors and how the classification accuracy performs under different segmentations of these areas. T1, T2, and T3 were used in the case of segmentation scenarios, while only T2 and T3 were used in the classification tests, due to their larger extent and contrasting landscapes: urban area (T2) and semi-natural landscape (T3) (Figure 1).

**Table 1.** Summary and characteristics of the three test areas.

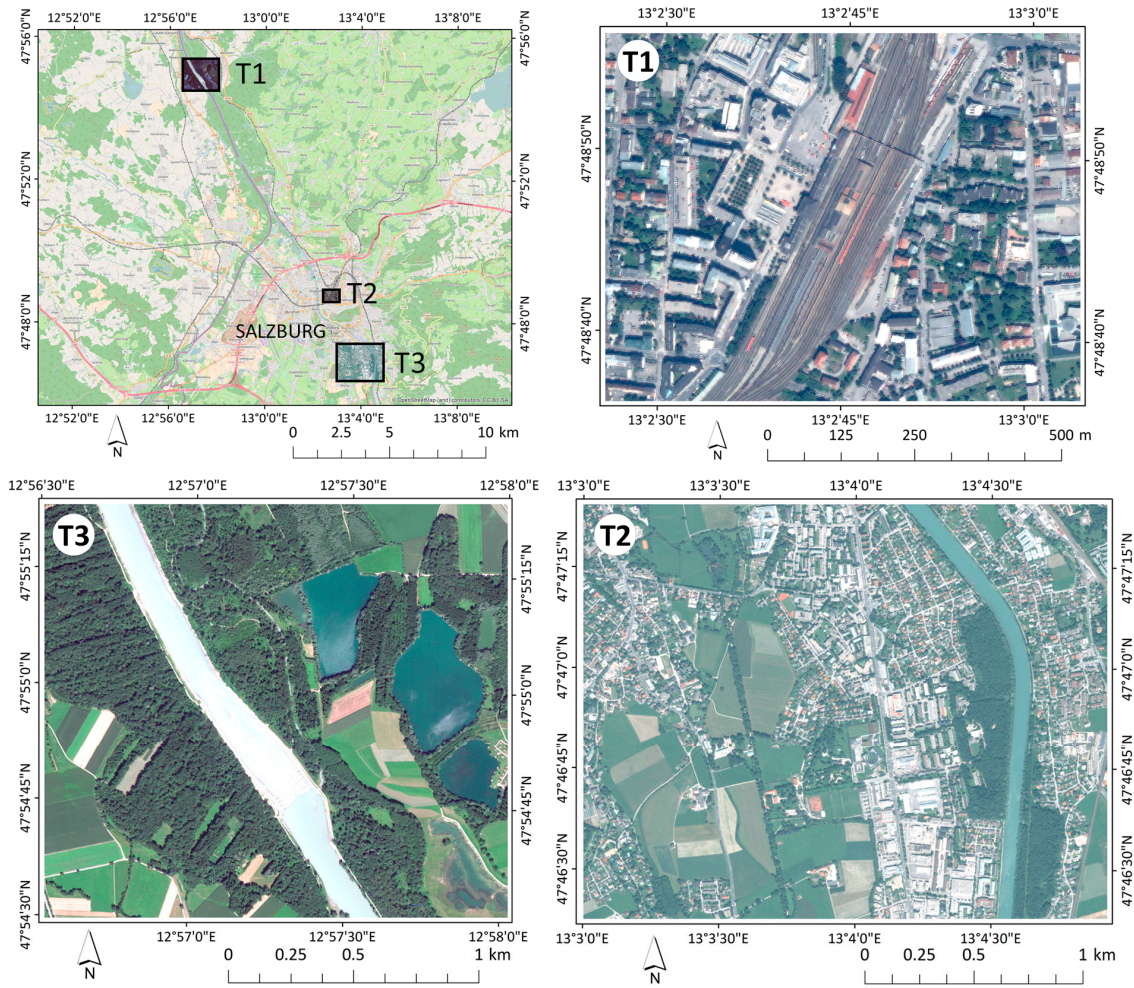| Test | Imagery | Spatial Resolution (m) | Number of Bands | Extent (pixels) | Length × Width (pixels) | Location |
|------|---------|------------------------|-----------------|-----------------|--------------------------|----------|
| T1 | QuickBird | 0.6 | 4 | 4,016,016 | 1347 × 1042 | City of Salzburg, Austria |
| T2 | QuickBird | 0.6 | 4 | 12,320,100 | 4004 × 3171 | City of Salzburg, Austria |
| T3 | WorldView-2 | 0.5 | 8 | 12,217,001 | 3701 × 3301 | 10 km north of city of Salzburg |

**Figure 1.** The three test areas under investigation (T1 to T3), as detailed in Table 1.

## 2.2. Simple Linear Iterative Clustering (SLIC) Superpixels

Various algorithms to compute superpixels exist, each with individual strengths and weaknesses [16]. In this study, we used Simple Linear Iterative Clustering (SLIC) [19,45] which has been proven to outperform other state-of-the-art superpixel methods [19,46], because of its simplicity, adherence to boundaries, computational speed, and memory efficiency [19]. SLIC has only one parameter, *k*, which is the desired number of equally sized superpixels to be generated. An optional parameter *m* can be set up, which controls the compactness of the superpixels [19]. The algorithm starts by generating *k* initial cluster centers on a regular grid spaced *S* pixels apart. To avoid placing the seed on an edge or a noisy pixel, cluster centers are perturbed in a $3 \times 3$ neighborhood to the lowest gradient position. An iterative procedure starts by assigning each pixel to the nearest cluster center using a distance measure *D* (Equation (1)), which combines distance of color proximity (Equation (2)) and distance of spatial proximity (Equation (3)):

$$D = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2} \tag{1}$$

$$d_c = \sqrt{\sum_{s_p \in B} \left(I\left(x_i, y_i, s_p\right) - I\left(x_j, y_j, s_p\right)\right)^2} \tag{2}$$

$$d_s = \sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2} \tag{3}$$

where $d_c$ and $d_s$ represent the color and spatial distance between pixels $I(x_i, y_i, s_p)$ and $I(x_j, y_j, s_p)$ in the spectral band $s_p$, $B$ represents the set of spectral bands used, $S$ is the sampling interval of the clusters centroids, and $m$ controls the compactness of superpixels [36]. The color distance controls superpixels homogeneity, while the spatial distance forces superpixel compactness [36].

SLIC is an adapted *k*-means clustering, but what makes it fast and computationally efficient is that SLIC does not compare each pixel with all pixels in the scene. For a region of approximate size $S \times S$, the distance $D$ is computed in a region $2S \times 2S$ around the superpixel center, reducing the number of $D$ calculations. After each pixel is assigned to the nearest cluster center, the new cluster centers are updated for the newly generated superpixels, representing the mean vector of all the pixels belonging to the superpixel. A residual error $E$ is computed, as the distance between previous center locations and recomputed center locations. The iterative procedure stops when the error converges or when it reaches the threshold of input iterations. Achanta, et al. [19] found out that 10 iterations are sufficient for most images and, therefore, we used this value for SLIC superpixel generation. In the end, a post processing step enforces connectivity by reassigning disjoint pixels to nearby superpixels [19].

The same authors proposed a parameter-free SLIC version (SLICO), which generates regular shaped superpixels across the scene, regardless of textured or non-textured regions in the image. On the contrary, SLIC is sensitive to texture, generating smooth regular-sized superpixels in non-textured regions and highly irregular superpixels in textured regions [19].

For generating superpixels, we used a freely available GDAL implementation (GDAL-segment), available on https://github.com/cbalint13/gdal-segment. The tool needs the following parameters: input raster image, output shapefile of superpixel polygons, the algorithm of generating superpixels, the number of iterations (the default is 10) and the size of the superpixels to be generated (the default is 10).

For segmentation tests, superpixels were generated using SLIC and SLICO algorithms [19] with an initial size of $10 \times 10$ pixels. We used this value after trial-and-error, in order to avoid an extreme oversegmentation (and thus going back to the pixel level) or too large superpixels (which contain more than one class). No multiple generations of SLIC superpixels with the same size were needed to test the stability of the method, since the algorithm gives the same delineation of superpixels each time. For classification tests, we used the same SLIC and SLICO algorithms for generating four sizes of superpixels: $5 \times 5$, $10 \times 10$, $15 \times 15$, and $20 \times 20$ pixels, in order to evaluate the sensitivity of classification accuracy to the size of the superpixels.

### 2.3. Multiresolution Segmentation: Pixels vs. Superpixels

A common algorithm in OBIA for segmenting an image into image objects is multiresolution segmentation (MRS) [14]. It starts from pixel level and successively aggregates them into objects of different shapes, sizes, and characteristics, until it reaches a threshold of homogeneity set by the user. One of the biggest issues in MRS is the selection of parameters, of which the most important one is the scale parameter (SP). Advancements were achieved in objectively select the SPs for MRS [6,47,48] and thus aiming for an automated MRS. Drăguţ, et al. [47] developed Estimation of Scale Parameter tool (ESP) to detect optimal scales based on a local variance graph, using a single layer. Drăguţ, et al. [6] extended this approach into an automated tool (ESP2) for multiple layers. The ESP2 tool is a fully automated methodology for the selection of scale parameters to extract three distinct scales using MRS, implemented in the eCognition Developer software (Trimble Geospatial).

For reasons of objectivity, we used the ESP2 tool starting from the pixel level, as usual, and starting from the superpixel level, using a hierarchical bottom-up region merging approach (i.e., starting from an initial level, the next level is generated based on the previous one) to derive only the finest level of objects (Level 1 of the hierarchy approach of ESP2) (Figure 2).

### 2.4. Assessment of Segmentation Results

The segmentations results were evaluated by comparing the geometries of resulted objects with reference objects (Figure 2). We independently manually delineated 50 reference objects for each test area, representing very prominent features in the image, like houses, small buildings, groups of trees, singular trees, small agricultural fields, water bodies, and roads. There are a variety of methods to evaluate the segmentation results [49]. In this study, we used five measures of accuracy: Area fit index (AFI), Over-segmentation (OS), Under-segmentation (US), Root mean square (D), and Quality rate (QR), with a minimum percent overlap of 50% (Table 2). In the case of a perfect match between the geometries of objects, AFI, OS, US, and D would be 0 and QR would be 1. The measures are implemented into an eCognition tool by Eisank, et al. [50].

Besides geometry comparison, we measured the computational time needed for both approaches. For the ESP2 starting from pixel level we retained only the time needed to run the tool, while for ESP2 starting from superpixel level, we added to the time ESP2 needs to run the computational time for generating superpixels.
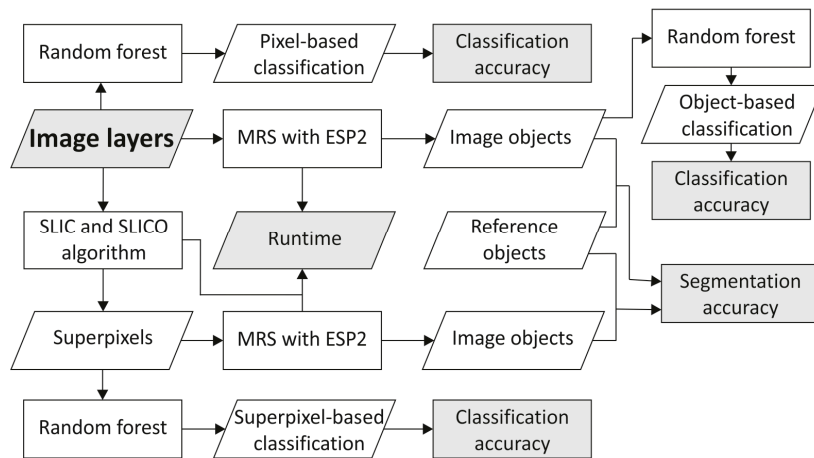


**Figure 2.** Proposed workflow to partition a scene into objects starting from pixels and superpixel level, respectively, by applying Estimation of Scale Parameter 2 (ESP2) [6], an automated parameterization tool for multiresolution segmentation (MRS) [14]. The computational time is retained for both cases. Resulted objects were geometrically compared with the reference objects and segmentation accuracy measures were derived. For the classification scenarios, the final accuracies were compared between pixel-based, object-based and superpixel-based classification.

**Table 2.** Overview of the selected segmentation accuracy measures (*C* indicates the total area of evaluated objects; *R* indicates the total area of reference objects).

| Measure | Equation | Domain | Ideal Value | Authors |
|---|---|---|---|---|
| Over-segmentation | $OS = 1 - \frac{C \cap R}{R}$ | [0, 1] | 0 | Clinton, et al. [51] |
| Under-segmentation | $US = 1 - \frac{C \cap R}{C}$ | [0, 1] | 0 | Clinton, et al. [51] |
| Area fit index | $AFI = \frac{R - C}{R}$ | oversegmentation: AFI > 0 undersegmentation: AFI < 0 | 0 | Lucieer and Stein [52] |
| Root mean square | $D = \sqrt{\frac{OS^2 + US^2}{2}}$ | [0, 1] | 0 | Clinton, et al. [51] |
| Quality rate | $QR = 1 - \frac{C \cap R}{C \cup R}$ | [0, 1] | 1 | Winter [53] |

### 2.5. Training and Validation Samples

For both test areas used in the classification process, T2 and T3, five classes of interest were identified: build-up area, woodland, grassland, bareland, and water for T2 and woodland, grassland, bareland, lake, and river for T3 (Table 3). In order to create training and validation samples, a number of 2000 points was randomly generated for both test areas. Out of them, 1078 samples

for T2 and 1225 samples for T3 were visually labeled into one of the five classes, following the class representativeness across the scenes (Figure 3). Furthermore, we split the datasets into 50% training samples and 50% validation samples, resulting in 540 training samples and 538 validation samples for T2 and 614 training samples and 611 validation samples for T3 (Table 3). For the superpixel-based and object-based classification, the training samples were represented by the superpixels/objects which had at least one training point sample inside its borders. The validation was done based on points.

**Table 3.** Training and validation samples for T2 and T3 test areas.

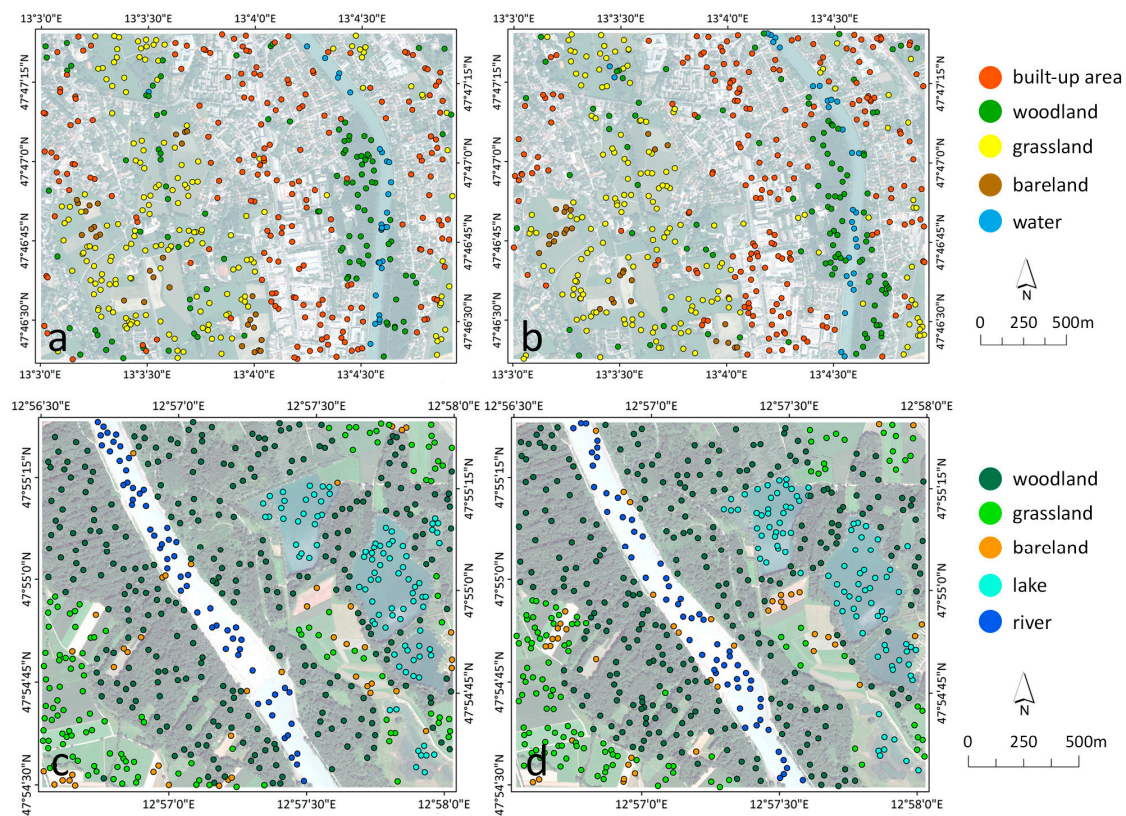| | T2—QuickBird | | | T3—WorldView-2 | |
|---|---|---|---|---|---|
| | **Training** | **Validation** | | **Training** | **Validation** |
| Build-up area | 199 | 199 | Woodland | 297 | 296 |
| Woodland | 107 | 107 | Grassland | 121 | 120 |
| Grassland | 174 | 173 | Bareland | 44 | 44 |
| Bareland | 30 | 29 | Lake | 59 | 58 |
| Water | 30 | 30 | River | 93 | 93 |
| Total | 540 | 538 | Total | 614 | 611 |



**Figure 3.** Spatial distribution of samples used for classification: (**a**) training and (**b**) validation samples for QuickBird and (**c**) training and (**d**) validation samples for WorldView-2 datasets.

*2.6. Random Forest Classification*

Random forest (RF) classifier, an ensemble of classification trees [54], was used in order to classify the T2 and T3 scenes. A detailed overview of the RF is given in [54], while details about recent developments of RF in remote sensing can be found in [55]. Briefly, RF builds a set of trees that are created by selecting a subset of training samples through a bagging approach, while the remaining samples are used for an internal cross-validation to estimate how well the RF model performs [55]. The final membership class is represented by the one with the maximum votes from the user-defined

number of trees (*Ntree*) used to grow the forest. In most of the studies reported in [55], *Ntree* value of 500 proved sufficient in order to stabilize the errors before reaching this number of classification trees.

**Table 4.** Overview of object attributes used in random forest (RF), as exported from the eCognition Developer [56].

| Type | Variable | Definition [56] |
|------|----------|-----------------|
| Spectral | Mean band x<br>Standard deviation band x<br>Brightness<br>NDVI | The mean layer x intensity value of an object/pixel<br>The standard deviation of an object/pixel in band x<br>The mean value of all the layers used for RF<br>Normalized Difference Vegetation Index |
| Texture | GLCM standard deviation<br>GLCM homogeneity<br>GLCM correlation | Gray level co-occurrence matrix (GLCM) [57] |
| Shape | Border index | The ration between the border lengths of the object and the smallest enclosing rectangle |
| | Compactness | The product of the length and width, divided by the number of pixels |
| | Area | The area (in pixels) of an object |

The "randomForest" R package was used [58,59], with the default value for *Ntree* of 500 and using spectral, shape and texture attributes, as detailed in Table 4. In the case of superpixel-based and object-based RF classifications, a number of 16 object attributes was used in the case of QuickBird image (4 bands) and 24 object attributes for WorldView-2 image (8 bands). For the pixel-based RF classification, only the spectral attributes were used (10 for QuickBird and 18 for WorldView-2, respectively), since no shape or texture information is enclosed inside a single pixel.

*2.7. Classification Accuracy Evaluation*

The error matrix was used in order to extract indicators to summarize specific information about classification performance [60,61]. Overall accuracy (OA) represents the ratio between the total correctly classified pixels and the total pixels in the error matrix [61]. Producer's accuracy (PA) indicates the probability that a pixel in the reference classification have the same class in the classified image, while the User's accuracy (UA) represents the probability that a pixel in the classified image represents the same class in the reference classification [60,61]. The two measures are obtained by dividing the total correct by the row total (UA) or by the column total (PA) [61]. Kappa coefficient [62] indicates that the results presented in the error matrix are significantly better than random results, based on the difference between the actual agreement in the error matrix and the chance agreement which is indicated by the total, specific of rows and columns [61,63].

**3. Results**

*3.1. SLIC and SLICO Superpixel Generation*

All tests were conducted on a computer station with Intel Core i5-4590 CPU (3.30 GHz) processor with 8 GB RAM, using a 64-bit Windows 7 operating system.

Generating SLIC and SLICO superpixels is a very fast and memory efficient procedure. In the case of SLIC superpixels of initial size of $10 \times 10$ pixels, the scored runtime was 2 s for T1, 18 s for T2, and 26 s for T3, while for SLICO was 3 s for T1, 34 s for T2, and 37 s for T3. Computing SLICO superpixels is a slightly slower procedure because SLICO generates regular shaped superpixels across the scene, thus having supplementary computational constraints (i.e., compactness) (Figure 4). Test areas T2 and T3 have similar extents but different number of bands (4 and 8, respectively) which influence the runtime in favor of the former. However, because T3 is of lower landscape complexity than T2, the

runtime difference is not very big in the case of SLIC superpixels (8 s) and almost negligible for SLICO superpixels (3 s).

To evaluate how the computational speed increases with the increasing number of pixels, we derived SLIC superpixels for a 4-band QuickBird scene with extents ranging from 1 to 10 million pixels, with an increment of 1 million (Figure 5). From the resulted runtime, which ranged between 2 s and 16 s, we can conclude that for a 4-band VHR satellite image, with each layer of one million pixels, the SLIC algorithm needs approx. 2 s to extract superpixels (i.e., for a 30 million pixels QuickBird scene it will need approx. 1 min to generate SLIC superpixels).
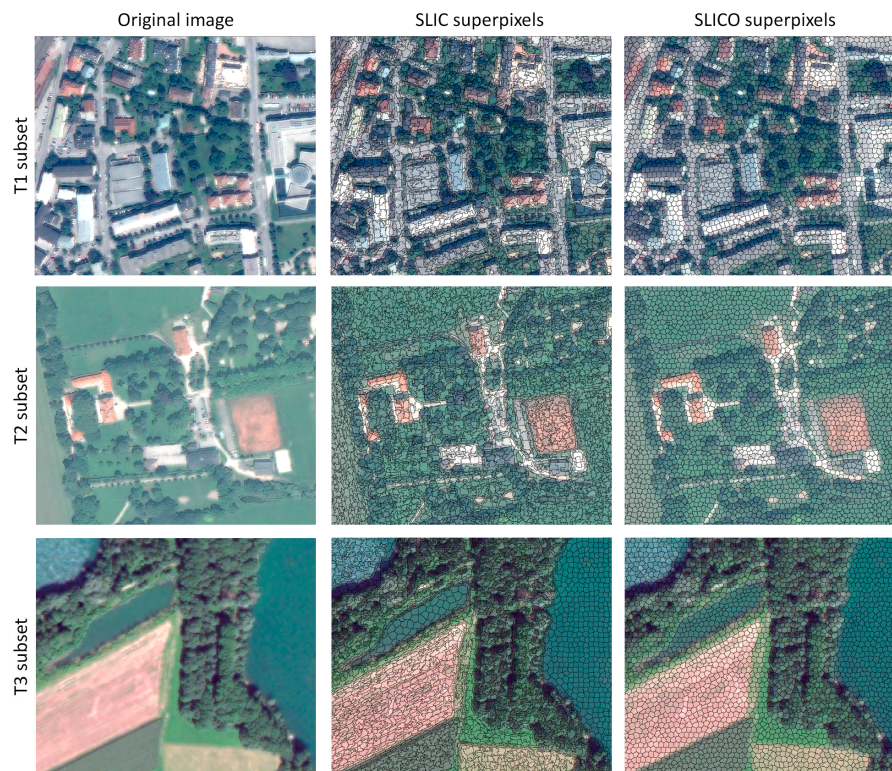


**Figure 4.** Comparison of Simple Linear Iterative Clustering (SLIC) and SLICO superpixel adherence to natural image boundaries derived using initial clustering of $10 \times 10$ pixels.
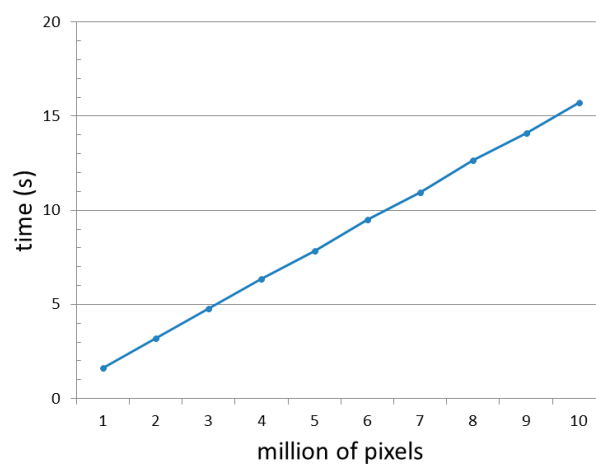


**Figure 5.** Runtime (in seconds, vertical axis) comparison for generating SLIC superpixels for a QuickBird image with 4 bands and extent ranging from 1 to 10 million pixels, with an increment of 1 (horizontal axis).

*3.2. Multiresolution Segmentation: Pixels* vs. *Superpixel Results*

　　The complexity of the three test areas was reduced by a few orders of magnitude, from 1.4 mil pixels to 12.835 SLIC superpixels for T1, from 12.7 mil pixels to 123.153 SLIC superpixels for T2 and from 12.2 mil pixels to 131.415 SLIC superpixels for T3, respectively (Table 5). Due to this fact, SLIC and SLICO superpixels make a big difference in any subsequent processing regarding the computational time. The larger the scene and the bigger the number of bands, the bigger the time needed for ESP2 to run starting from a pixel-grid (from 1 min 29 s to 5 h 35 min 24 s) (Table 5). Runtime of ESP2 starting from SLIC superpixels is higher than that of ESP2 starting from SLICO superpixels, the difference ranging between 3 s for T1 and 3 min 13 s for T2. This is explained by the fact that the scale parameter detected for the latter case is smaller and therefore the processing of ESP2 stops earlier. This is justified by the fact that SLICO superpixels have compactness constraints to follow a regular lattice and, as a consequence, they can omit meaningful image boundaries, increasing the internal heterogeneity of the superpixels. SLIC superpixels reduced the runtime by a magnitude of 229% for T1, 1146% for T2 and 2476% for T3, respectively.

**Table 5.** Segmentation accuracy and computational time for the three test areas (T1 to T3) for multiresolution segmentation (MRS) using pixels, Simple Linear Iterative Clustering (SLIC), and SLICO superpixels, respectively. The number of initial pixels/superpixels, scale parameter (SP) and the final number of objects after applying ESP2 is depicted in the table. The segmentation accuracy measures used are Area fit index (AFI), Over-segmentation (OS), Under-Segmentation (US), Root mean square (D) and Quality rate (QR).

| Test | | Segmentation Results | | | Segmentation Accuracy Metrics | | | | | Time |
|------|--------|-----------|-----|-------------------|-------|-------|-------|-------|-------|----------------|
| | | Number | SP | Number of Objects | AFI | OS | US | D | QR | |
| **T1** | Pixels | 1,403,574 | 69 | 3,109 | 0.499 | 0.560 | 0.121 | 0.405 | 0.414 | 1 min 29 s |
| | SLIC | 13,835 | 81 | 2,017 | 0.388 | 0.463 | 0.122 | 0.338 | 0.499 | 27 s |
| | SLICO | 13,906 | 61 | 2,757 | 0.447 | 0.515 | 0.122 | 0.374 | 0.454 | 24 s |
| **T2** | Pixels | 12,696,684 | 172 | 4,670 | 0.174 | 0.229 | 0.067 | 0.169 | 0.729 | 2 h 42 min 40 s |
| | SLIC | 123,153 | 173 | 4,204 | 0.088 | 0.161 | 0.079 | 0.127 | 0.782 | 13 min 02 s |
| | SLICO | 125,842 | 148 | 5,354 | 0.335 | 0.386 | 0.075 | 0.278 | 0.584 | 9 min 49 s |
| **T3** | Pixels | 12,217,001 | 220 | 1,632 | 0.100 | 0.148 | 0.052 | 0.111 | 0.813 | 5 h 35 min 24 s |
| | SLIC | 131,415 | 212 | 1,702 | 0.062 | 0.124 | 0.066 | 0.099 | 0.823 | 13 min 03 s |
| | SLICO | 121,525 | 172 | 2,338 | 0.223 | 0.275 | 0.066 | 0.200 | 0.688 | 10 min 46 s |

　　The scale parameters for pixel-grid and SLIC superpixels are similar. The most evident case is for T3 (220 and 212, respectively), where approximately the same number of objects was extracted in the end (1632 and 1702, respectively).

　　Compared with the other two approaches, SLIC superpixels had better values for AFI, OS, D, and QR, and slightly worse values for US than the pixel approach (the difference range from 0.001 for T1 and 0.014 for T3) (Table 5) (Figure 6). In the latter case, this can be considered a negligible difference when aiming to reduce the runtime. Compared to the pixel approach, SLIC had better values of QR for T1 (0.499 compared to 0.414), for T2 (0.782 compared to 0.729), and for T3 (0.823 compared to 0.813). The oversegmentation of the scene using SLICO superpixels negatively impacts the accuracy measures and, therefore, they are better than the pixel approach only in the case of T1 (QR of 0.454 compared to 0.414), where the number of objects is slightly smaller. In test area T1, even if there is a big difference in the number of objects extracted between pixel approach and SLIC superpixels (3109 and 2017, respectively), the US has the same value, while OS decreases by 0.1. This means that SLIC superpixels are better at adhering to the boundaries and, therefore, creating more meaningful objects than those generated by starting from a pixel-grid.
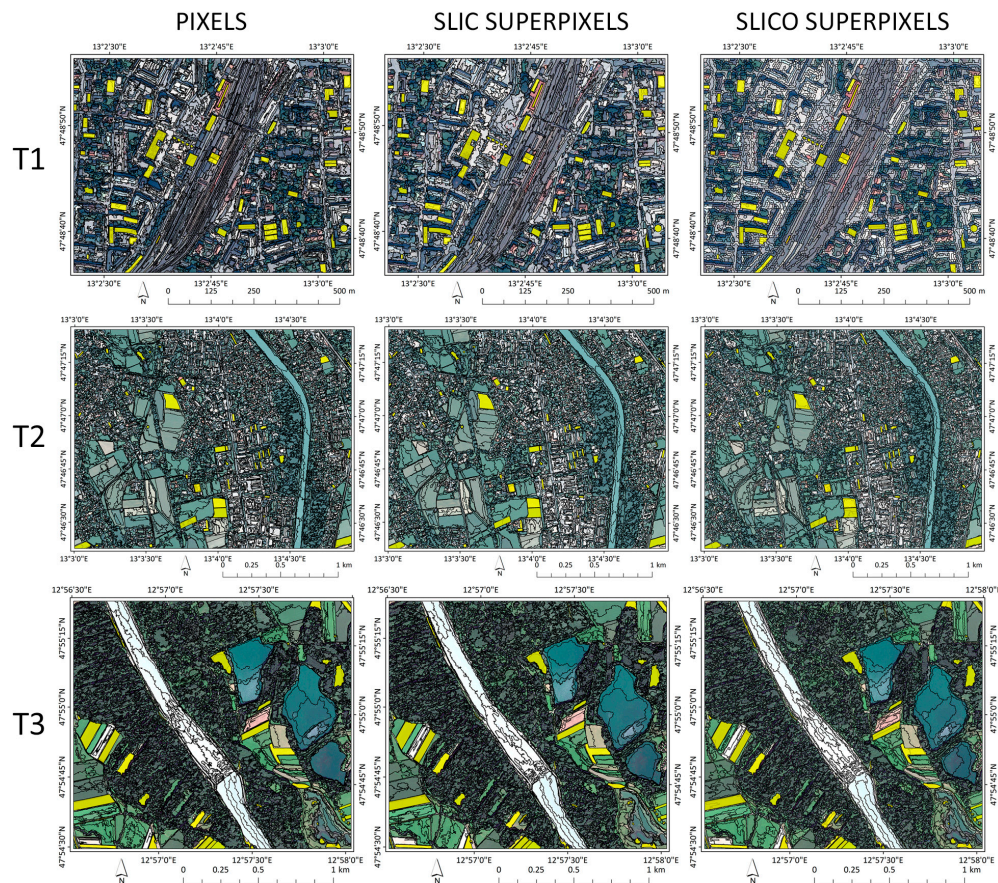
**Figure 6.** Segmentation results (black outline) and reference polygons (solid yellow) for all test areas (horizontal) and for all three cases of segmentation (vertical).

### 3.3. Pixel-Based, Superpixel-Based and MRS-Based RF Classification Results

Ten variations of classifications were compared for both test areas: one pixel-based classification and nine object-based classifications (one case with objects derived from automated MRS starting from pixel level and eight cases represented by four different sizes of SLIC and SLICO superpixels, respectively).

In the case of T2 area, all the classifications kept the general pattern of class distributions within the image (Figure 7). However, a quantitative analysis reveals differences between approaches. Pixel-based classification has the lowest OA (91.45%) and this is due to the confusion between bareland and build-up areas, giving the lowest values of PA for build-up areas (91.46%) and lowest UA for bareland class (56.62%). Also, the pixel-based approach has difficulties in distinguishing woodland from grassland, having the lowest PA for woodland (85.98%) and the lowest UA for grassland (91.62%) (Figure 8). Unexpectedly, classified objects resulted from MRS (OA of 96.09%) and SLIC and SLICO superpixels have the classification accuracy comparable (OA between 95.17% and 97.21%) (Table 6). Even more, all the four sizes of SLIC superpixels and two SLICO superpixels (5 × 5 and 15 × 15) outperformed the MRS-based classification regarding the OA values. The highest OA was achieved by SLIC 20 × 20 RF classification, of 97.21%.

In the case of T3 area, the results follow the same trend as in the T2 area, meaning that the spatial structure of class distributions is clearly visible in all classification scenarios (Figure 9). Being a less heterogeneous scene, comprising natural elements, the OA values are higher than in T2 classification evaluation: more than 99% for MRS-based classification and all eight variations of superpixel-based classification. The highest OA was achieved by SLIC 15 × 15 superpixels (99.84%), while the MRS-based classification reached a value of 99.51% for OA (Table 6). In the case of pixel-based RF classification,

lowest OA (90.51%) is due to the confusions between woodland and grassland, which gives the lowest PA for woodland (81.08%) and lowest UA for grassland (67.82%) (Figure 10).
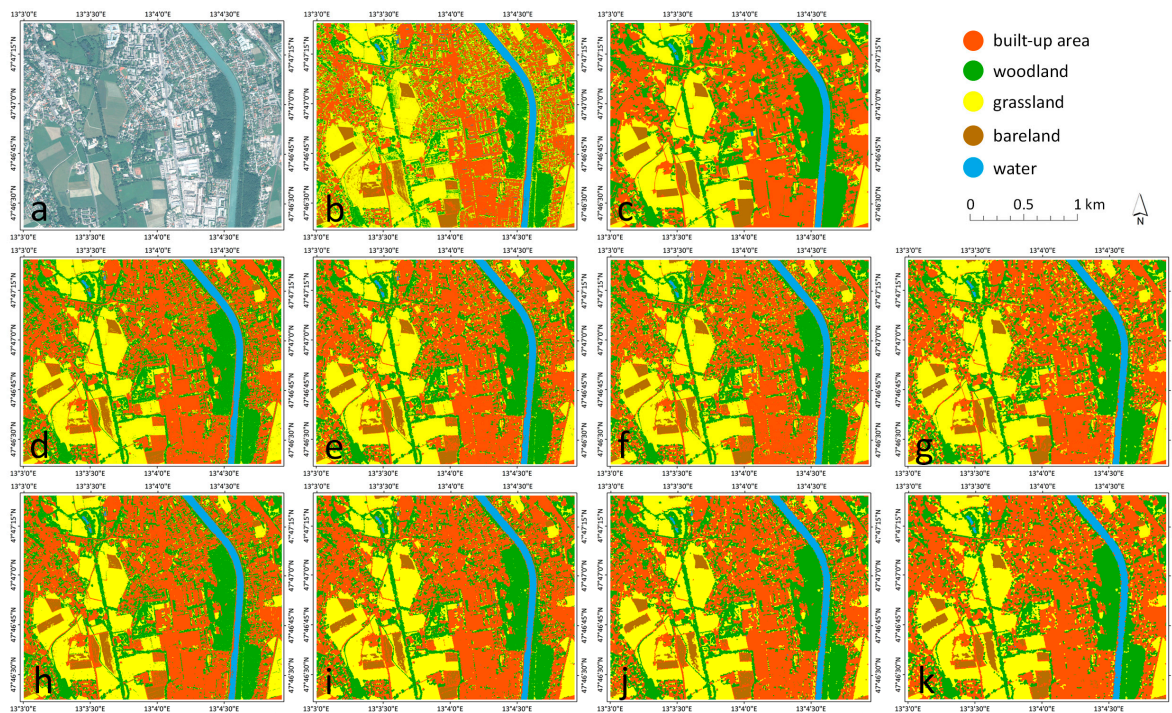


**Figure 7.** Random forest (RF) classification results for T2: (**a**) original QuickBird RGB image, (**b**) pixel-based classification, (**c**) multiresolution segmentation classification, (**d**–**g**) SLIC superpixels classification of $5 \times 5$, $10 \times 10$, $15 \times 15$ and $20 \times 20$ pixels initial size of superpixels and (**h**–**k**) SLICO superpixels classification of $5 \times 5$, $10 \times 10$, $15 \times 15$ and $20 \times 20$ pixels initial size of superpixels.
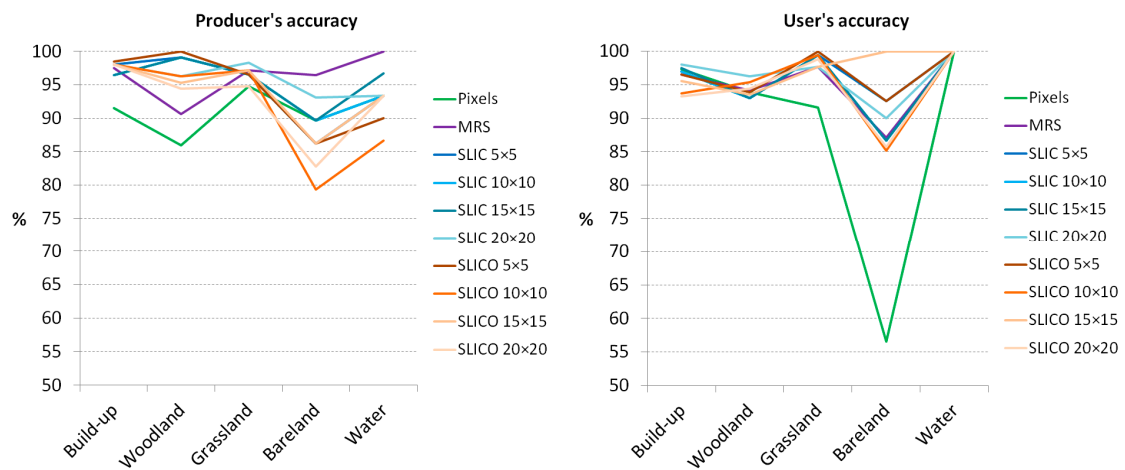


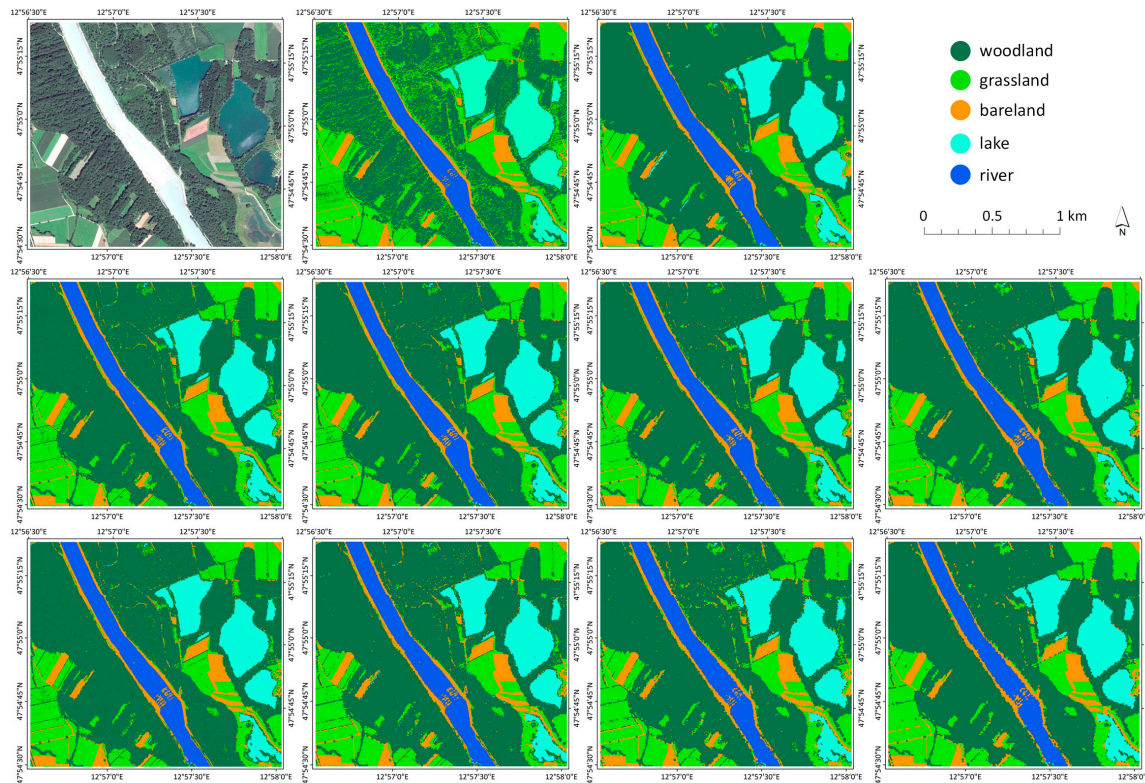**Figure 8.** Producer's and user's accuracy for T2 RF classification results.

**Figure 9.** Random forest classification results for T3: (**a**) original WorldView-2 RGB image, (**b**) pixel-based classification, (**c**) multiresolution segmentation classification, (**d**–**g**) SLIC superpixels classification of 5 × 5, 10 × 10, 15 × 15 and 20 × 20 pixels initial size of superpixels and (**h**–**k**) SLICO superpixels classification of 5 × 5, 10 × 10, 15 × 15 and 20 × 20 pixels initial size of superpixels.
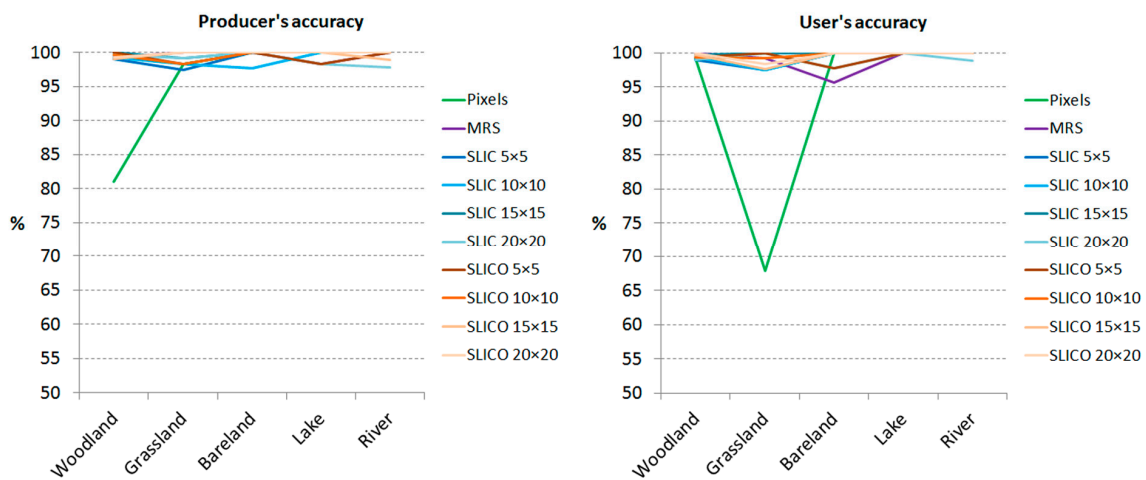


**Figure 10.** Producer's and user's accuracy for T3 RF classification results.

**Table 6.** Overall accuracy (OA) and Kappa index for T2 and T3 RF classifications, using all ten approaches tested in this study.

|  | T2-QuickBird | | T3-WorldView-2 | |
| --- | --- | --- | --- | --- |
|  | OA (%) | Kappa | OA (%) | Kappa |
| Pixels | 91.45 | 0.881 | 90.51 | 0.867 |
| MRS | 96.09 | 0.945 | 99.51 | 0.993 |
| SLIC 5 × 5 | 96.84 | 0.956 | 99.02 | 0.956 |
| SLIC 10 × 10 | 96.47 | 0.951 | 99.18 | 0.988 |
| SLIC 15 × 15 | 96.65 | 0.953 | 99.84 | 0.998 |
| SLIC 20 × 20 | 97.21 | 0.961 | 99.18 | 0.998 |
| SLICO 5 × 5 | 97.03 | 0.958 | 99.51 | 0.988 |
| SLICO 10 × 10 | 95.72 | 0.940 | 99.51 | 0.993 |
| SLICO 15 × 15 | 96.28 | 0.948 | 99.35 | 0.991 |
| SLICO 20 × 20 | 95.17 | 0.932 | 99.67 | 0.995 |

## 4. Discussion

The novelty of this study consists in using SLIC superpixels to decompose the complexity of VHR remote sensing data into perceptually meaningful building blocks to: (1) significantly decrease the time needed for automated MRS and (2) create fast and accurate thematic maps of classes of interest.

In the case of superpixel-based MRS segmentation, the computational time decreased from a couple of hours to a few minutes, while maintaining or improving the geometric accuracy of extracted objects. Several consequences emerge from this, such as increasing the maximum extents for running ESP2 Supplementary tests (not shown here) proved that for scenes of tens of millions of pixels ESP2 is successfully fast when starting from SLIC superpixels, while in the case of starting from pixels it crashes due to the immense amount of resources needed to compute the statistics at the pixel-level. An unexpected finding is related to the fact that the geometric accuracy of our proposed workflow is at least similar or slightly better than MRS starting from pixel-level. This is directly linked to how the two segmentation algorithms, SLIC and MRS, perform the initial grouping of pixels, with a better adherence to image objects boundaries in the case of the former. This fact needs to be further investigated in our future studies.

Comparing pixel-based, object-based and superpixel-based classification revealed the fact that we can achieve a similar or even better classification accuracy using superpixels instead of objects resulting from automated MRS. Gao, et al. [64] underlined that the best classification accuracy can be obtained by using an optimal segmentation, avoiding over- or under-segmentation of the image. However, Belgiu and Drăguţ [65] found that classification accuracy is significantly less dependent on the segmentation results and as long as under-segmentation remains at acceptable levels, we can still achieve a high classification accuracy. In our case, although we used four different sizes of superpixels (5 × 5, 10 × 10, 15 × 15 and 20 × 20, respectively), we achieved very similar OA, between 95.15% and 97.21% for T2 and between 99.02% and 99.84% for T3. Therefore, even with these over-segmentations generated by SLIC and SLICO algorithms we can achieve similar results to the object-based classification, with objects obtained by applying a time-consuming automated MRS.

The object variables with the highest importance in RF classification were the spectral features, while the texture and shape features shared the least importance roles in the classification process. In the case of superpixels, the shape features play an insignificant role because most of the superpixels have similar sizes and shapes.

We used SLIC and SLICO superpixels because they were shown to outperform other state-of-the-art superpixel algorithms [19,46]. In our study, applying MRS on SLIC superpixels proved to be the most efficient in terms of geometric accuracy of the final objects. Even if SLICO superpixels gave better geometric accuracy results than the pixel-approach only in the case of T1, we included them because SLICO can be a valid option in studies were a more regular lattice is needed, where

the complexity of the scene is lower or when the features of interest are widely larger than the size of SLICO superpixels. Another notable advantage is that most used superpixel algorithms are available in open source implementations.

Besides many advantages, the drawbacks of using superpixels are (1) the choice of the superpixel algorithm and its parameters, (2) choosing between a lattice or non-lattice arrangement of the superpixels, and maybe the most important one, (3) the risk of losing meaningful image boundaries by placing them inside a superpixel [16], if the superpixel size is larger than the size of objects of interest. The size of the superpixels needs to be carefully chosen, so as not to worsen the computational efficiency (by generating too small superpixels) or to contain more than one class inside a superpixel (by having too coarse superpixels). Usage of smaller superpixels can increase the geometric accuracy of the objects, but it will influence the computational time, while larger superpixels would omit important boundaries and increase internal heterogeneity of the objects (Figure 11). After testing different sizes of the SLIC and SLICO superpixels in MRS, we suggest that using an initial size of the superpixel of $10 \times 10$ is a good compromise to enable the advantages of using superpixels, both in terms of accuracy of the final results and computational runtime of ESP2. Another minor drawback of using SLIC and SLICO superpixels in MRS or for a thematic classification is that we introduce further parameterization in the process, but the only parameter one has to set is the desired size of the generated superpixels, that means how finer or coarser the generated superpixels should be.
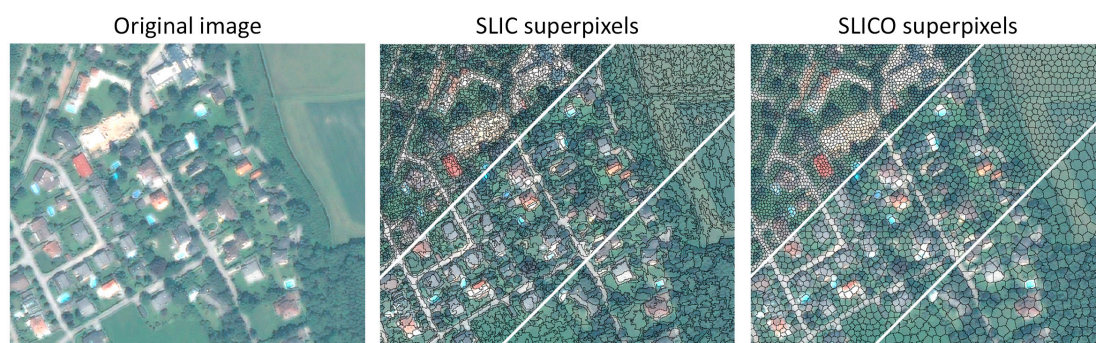


**Figure 11.** Visual comparison of superpixels produced by SLIC and SLICO methods, on a subset of T2. The average superpixel size in the upper left of the image is $5 \times 5$ pixels, in the middle of the image is $10 \times 10$ pixels and $15 \times 15$ pixels in the lower right.

Using VHR images with different characteristics gave us an overview of their effects in generating superpixels. For the same size of desired superpixels, the larger the extent of the scene and the higher the number of layers, the higher the computational time. However, the most important influence over the runtime is the extent of the generated superpixels: smaller superpixels will require longer time than larger superpixels. Also, a higher number of iterations in the clustering process will increase the runtime. As we already mention, 10 iterations were found to be sufficient for most of the images [19]. A higher number of iterations can lead to a better adherence of the superpixels to the boundaries, but it will affect the time needed for generating them.

Superpixels are used in many computer vision applications, like object recognition [66], image segmentation [67], depth estimation [68] and object localization [69] as well as biomedical applications [70]. In remote sensing, superpixels can help in overcoming the current limitations regarding the computational time and memory efficiency of studies dealing with image segmentation or thematic classifications of large remote sensing data. Furthermore, superpixels have already been extended and applied for the analysis of 3D scene structure [71]. The 3D version of superpixels—supervoxels—has shown robustness and computational speed when having to face billions of voxels from 3D electron microscope image stacks [70]. Transferring this into the spatial domain, superpixels and supervoxels are worth testing in 3D analysis, such as geological modelling,

environment modelling or applications dealing with terrestrial or airborne laser scanning data, just to name a few. Superpixels also receive high attention and are of paramount importance in video processing applications [72] and since we now capture HD videos from satellites and unmanned aerial vehicle (UAV), we can aim at live processing of these satellite remote sensing videos for different purposes. Thus, we can further expand the capabilities of remote sensing applications to explain and to solve current problems of humankind.

## 5. Conclusions

In this paper, we showed how to efficiently partition an image into objects by using SLIC superpixels as the starting point for MRS and how the SLIC superpixels could be used for fast and accurate thematic mapping. From the segmentation point of view, when compared to the traditional approach (starting from pixel-grid), our approach outperformed both in terms of geometric accuracy of the extracted objects and computational time. From the classification point of view, SLIC superpixels successfully replaced the objects obtained from MRS in a RF classification, achieving similar or better overall accuracy of the final thematic maps. These two approaches have the potential to enhance the automation of big remote sensing data analysis, processing, and labelling, especially when time is an important constraint.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Hay, G.J.; Castilla, G. Geographic object-based image analysis (geobia): A new name for a new discipline. In *Object-Based Image Analysis*; Blaschke, T., Lang, S., Hay, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 75–89.

2. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]

3. Cheng, G.; Han, J. A survey on object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 11–28. [CrossRef]

4. Neubert, M.; Herold, H.; Meinel, G. Evaluation of remote sensing image segmentation quality—Further results and concepts. In Proceedings of the International Conference on Object-Based Image Analysis (ICOIA), Salzburg University, Salzburg, Austria, 4–5 July 2006.

5. Arvor, D.; Durieux, L.; Andrés, S.; Laporte, M.-A. Advances in geographic object-based image analysis with ontologies: A review of main contributions and limitations from a remote sensing perspective. *ISPRS J. Photogramm. Remote Sens.* **2013**, *82*, 125–137. [CrossRef]

6. Drăguţ, L.; Csillik, O.; Eisank, C.; Tiede, D. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 119–127. [CrossRef] [PubMed]

7. Benz, U.C.; Hofmann, P.; Willhauck, G.; Lingenfelder, I.; Heynen, M. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for gis-ready information. *ISPRS J. Photogramm. Remote Sens.* **2004**, *58*, 239–258. [CrossRef]

8. Liu, D.; Xia, F. Assessing object-based classification: Advantages and limitations. *Remote Sens. Lett.* **2010**, *1*, 187–194. [CrossRef]

9. Whiteside, T.G.; Boggs, G.S.; Maier, S.W. Comparing object-based and pixel-based classifications for mapping savannas. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 884–893. [CrossRef]

10. Ouyang, Z.-T.; Zhang, M.-Q.; Xie, X.; Shen, Q.; Guo, H.-Q.; Zhao, B. A comparison of pixel-based and object-oriented approaches to vhr imagery for mapping saltmarsh plants. *Ecol. Inf.* **2011**, *6*, 136–146. [CrossRef]

11. Im, J.; Jensen, J.R.; Tullis, J.A. Object-based change detection using correlation image analysis and image segmentation. *Int. J. Remote Sens.* **2008**, *29*, 399–423. [CrossRef]

12. Chen, G.; Hay, G.J.; Carvalho, L.M.T.; Wulder, M.A. Object-based change detection. *Int. J. Remote Sens.* **2012**, *33*, 4434–4457. [CrossRef]

13. Zhou, W.; Troy, A.; Grove, M. Object-based land cover classification and change analysis in the baltimore metropolitan area using multitemporal high resolution remote sensing data. *Sensors* **2008**, *8*, 1613–1636. [CrossRef] [PubMed]

14. Baatz, M.; Schäpe, A. Multiresolution segmentation-an optimization approach for high quality multi-scale image segmentation. In *Angewandte Geographische Informationsverarbeitung*; Strobl, J., Blaschke, T., Griesebner, G., Eds.; Wichmann-Verlag: Heidelberg, Germany, 2000; Volume 12, pp. 12–23.

15. Fisher, P. The pixel: A snare and a delusion. *Int. J. Remote Sens.* **1997**, *18*, 679–685. [CrossRef]

16. Neubert, P.; Protzel, P. Superpixel benchmark and comparison. In *Forum Bildverarbeitung 2012*; Karlsruher Instituts für Technologie (KIT) Scientific Publishing: Karlsruhe, Germany, 2012; pp. 1–12.

17. Ren, X.; Malik, J. Learning a classification model for segmentation. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Marseille, France, 13–16 October 2003; pp. 10–17.

18. Li, Z.; Chen, J. Superpixel segmentation using linear spectral clustering. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1356–1363.

19. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef] [PubMed]

20. Guangyun, Z.; Xiuping, J.; Jiankun, H. Superpixel-based graphical model for remote sensing image mapping. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5861–5871.

21. Shi, C.; Wang, L. Incorporating spatial information in spectral unmixing: A review. *Remote Sens. Environ.* **2014**, *149*, 70–87. [CrossRef]

22. Van den Bergh, M.; Boix, X.; Roig, G.; de Capitani, B.; Van Gool, L. Seeds: Superpixels extracted via energy-driven sampling. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 13–26.

23. Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905.

24. Felzenszwalb, P.F.; Huttenlocher, D.P. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **2004**, *59*, 167–181. [CrossRef]

25. Moore, A.P.; Prince, J.; Warrell, J.; Mohammed, U.; Jones, G. Superpixel lattices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.

26. Veksler, O.; Boykov, Y.; Mehrani, P. Superpixels and supervoxels in an energy optimization framework. In Proceedings of the eleventh European Conference on Computer Vision, Crete, Greece, 5–11 September 2010; pp. 211–224.

27. Comaniciu, D.; Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [CrossRef]

28. Vedaldi, A.; Soatto, S. Quick shift and kernel methods for mode seeking. In Proceedings of the European Conference on Computer Vision, Marseille, France, 12–18 October 2008; pp. 705–718.

29. Vincent, L.; Soille, P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 583–598. [CrossRef]

30. Levinshtein, A.; Stere, A.; Kutulakos, K.N.; Fleet, D.J.; Dickinson, S.J.; Siddiqi, K. Turbopixels: Fast superpixels using geometric flows. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2290–2297. [CrossRef] [PubMed]

31. Fourie, C.; Schoepfer, E. Data transformation functions for expanded search spaces in geographic sample supervised segment generation. *Remote Sens.* **2014**, *6*, 3791–3821. [CrossRef]

32. Ma, L.; Du, B.; Chen, H.; Soomro, N.Q. Region-of-interest detection via superpixel-to-pixel saliency analysis for remote sensing image. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1752–1756. [CrossRef]

33. Arisoy, S.; Kayabol, K. Mixture-based superpixel segmentation and classification of sar images. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1721–1725. [CrossRef]

34. Guo, J.; Zhou, X.; Li, J.; Plaza, A.; Prasad, S. Superpixel-based active learning and online feature importance learning for hyperspectral image analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 347–359. [CrossRef]

35. Li, S.; Lu, T.; Fang, L.; Jia, X.; Benediktsson, J.A. Probabilistic fusion of pixel-level and superpixel-level hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7416–7430. [CrossRef]

36. Ortiz Toro, C.; Gonzalo Martín, C.; García Pedrero, Á.; Menasalvas Ruiz, E. Superpixel-based roughness measure for multispectral satellite image segmentation. *Remote Sens.* **2015**, *7*, 14620–14645. [CrossRef]

37. Vargas, J.; Falcao, A.; dos Santos, J.; Esquerdo, J.; Coutinho, A.; Antunes, J. Contextual superpixel description for remote sensing image classification. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 1132–1135.

38. Garcia-Pedrero, A.; Gonzalo-Martin, C.; Fonseca-Luengo, D.; Lillo-Saavedra, M. A geobia methodology for fragmented agricultural landscapes. *Remote Sens.* **2015**, *7*, 767–787. [CrossRef]

39. Stefanski, J.; Mack, B.; Waske, B. Optimization of object-based image analysis with random forests for land cover mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2492–2504. [CrossRef]

40. Chen, J.; Dowman, I.; Li, S.; Li, Z.; Madden, M.; Mills, J.; Paparoditis, N.; Rottensteiner, F.; Sester, M.; Toth, C.; et al. Information from imagery: Isprs scientific vision and research agenda. *ISPRS J. Photogramm. Remote Sens.* **2016**, *115*, 3–21. [CrossRef]

41. Tiede, D.; Lang, S.; Füreder, P.; Hölbling, D.; Hoffmann, C.; Zeil, P. Automated damage indication for rapid geospatial reporting. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 933–942. [CrossRef]

42. Voigt, S.; Kemper, T.; Riedlinger, T.; Kiefl, R.; Scholte, K.; Mehl, H. Satellite image analysis for disaster and crisis-management support. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1520–1528. [CrossRef]

43. Lang, S.; Tiede, D.; Hölbling, D.; Füreder, P.; Zeil, P. Earth observation (eo)-based ex post assessment of internally displaced person (idp) camp evolution and population dynamics in zam zam, darfur. *Int. J. Remote Sens.* **2010**, *31*, 5709–5731. [CrossRef]

44. Strasser, T.; Lang, S. Object-based class modelling for multi-scale riparian forest habitat mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *37*, 29–37. [CrossRef]

45. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. *Slic Superpixels*; EPFL Technical Report 149300; School of Computer and Communication Sciences, Ecole Polytechnique Fedrale de Lausanne: Lausanne, Switzerland, 2010; pp. 1–15.

46. Csillik, O. Superpixels: The end of pixels in obia. A comparison of stat-of-the-art superpixel methods for remote sensing data. In Proceedings of the GEOBIA 2016: Solutions and Synergies, Enschede, The Netherlands, 14–16 September 2016; Kerle, N., Gerke, M., Lefevre, S., Eds.; University of Twente Faculty of Geo-Information and Earth Observation (ITC): Enschede, The Netherlands, 2016.

47. Drăguţ, L.; Tiede, D.; Levick, S. Esp: A tool to estimate scale parameters for multiresolution image segmentation of remotely sensed data. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 859–871. [CrossRef]

48. Kim, M.; Madden, M. Determination of optimal scale parameter for alliance-level forest classification of multispectral ikonos images. In Proceedings of the 1st International Conference on Object-based Image Analysis, Salzburg, Austria, 4–5 July 2006; Available online: http://www.isprs.org/proceedings/xxxvi/4-c42/papers/OBIA2006_Kim_Madden.pdf (accessed on 19 December 2016).

49. Zhang, Y.J. A survey on evaluation methods for image segmentation. *Pattern Recognit.* **1996**, *29*, 1335–1346. [CrossRef]

50. Eisank, C.; Smith, M.; Hillier, J. Assessment of multiresolution segmentation for delimiting drumlins in digital elevation models. *Geomorphology* **2014**, *214*, 452–464. [CrossRef] [PubMed]

51. Clinton, N.; Holt, A.; Scarborough, J.; Yan, L.I.; Gong, P. Accuracy assessment measures for object-based image segmentation goodness. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 289–299. [CrossRef]

52. Lucieer, A.; Stein, A. Existential uncertainty of spatial objects segmented from satellite sensor imagery. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2518–2521. [CrossRef]

53. Winter, S. Location similarity of regions. *ISPRS J. Photogramm. Remote Sens.* **2000**, *55*, 189–200. [CrossRef]

54. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

55. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

56. Trimble. *Ecognition Reference Book*; Trimble Germany GmbH: Munchen, Germany, 2012.

57. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. [CrossRef]

58. Liaw, A.; Wiener, M. Classification and regression by randomforest. *R News* **2002**, *2*, 18–22.

59. R Development Core Team. *R: A Language and Environment for Statistical Computing*; The R Foundation for Statistical Computing: Vienna, Austria, 2013.

60. Liu, C.; Frazier, P.; Kumar, L. Comparative assessment of the measures of thematic classification accuracy. *Remote Sens. Environ.* **2007**, *107*, 606–616. [CrossRef]

61. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [CrossRef]

62. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]

63. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; CRC: Boca Raton, FL, USA, 2009.

64. Gao, Y.; Mas, J.F.; Kerle, N.; Navarrete Pacheco, J.A. Optimal region growing segmentation and its effect on classification accuracy. *Int. J. Remote Sens.* **2011**, *32*, 3747–3763. [CrossRef]

65. Belgiu, M.; Drăguţ, L. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS J. Photogramm. Remote Sens.* **2014**, *96*, 67–75. [CrossRef] [PubMed]

66. Pantofaru, C.; Schmid, C.; Hebert, M. Object recognition by integrating multiple image segmentations. In *Computer Vision–ECCV 2008; Proceedings of the 10th European Conference on Computer Vision, Marseille, France, 12–18 October 2008*; Springer: Berlin, Heidelberg, Germany, 2008; pp. 481–494.

67. Li, Y.; Sun, J.; Tang, C.-K.; Shum, H.-Y. Lazy snapping. *ACM Trans. Graph. (ToG)* **2004**, *23*, 303–308. [CrossRef]

68. Zitnick, C.L.; Kang, S.B. Stereo for image-based rendering using image over-segmentation. *Int. J. Comput. Vis.* **2007**, *75*, 49–65. [CrossRef]

69. Fulkerson, B.; Vedaldi, A.; Soatto, S. Class segmentation and object localization with superpixel neighborhoods. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 27 September–4 October 2009; pp. 670–677.

70. Lucchi, A.; Smith, K.; Achanta, R.; Lepetit, V.; Fua, P. A fully automated approach to segmentation of irregularly shaped cellular structures in em images. In *Medical Image Computing and Computer-Assisted Intervention–Miccai 2010*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 463–471.

71. Saxena, A.; Sun, M.; Ng, A.Y. Make3d: Learning 3d scene structure from a single still image. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 824–840. [CrossRef] [PubMed]

72. Galasso, F.; Cipolla, R.; Schiele, B. Video segmentation with superpixels. In *Computer Vision–ACCV 2012; 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5–9, 2012*; Springer: Berlin Heidelberg, Germany, 2012; pp. 760–774.