remote sensing

MDPI

Article

Modified Superpixel Segmentation for Digital Surface Model Refinement and Building Extraction from Satellite Stereo Imagery

Zeinab Gharibbafghi *, Jiaojiao Tian and Peter Reinartz

German Aerospace Center (DLR), Remote Sensing Technology Institute, 82234 Wessling, Germany; jiaojiao.tian@dlr.de (J.T.); peter.reinartz@dlr.de (P.R.)

* Correspondence: zeinab.gharibbafghi@dlr.de; Tel.: +49-8153-28-4273

Received: 15 October 2018; Accepted: 15 November 2018; Published: 17 November 2018

Abstract: Superpixels, as a state-of-the-art segmentation paradigm, have recently been widely used in computer vision and pattern recognition. Despite the effectiveness of these algorithms, there are still many limitations and challenges dealing with Very High-Resolution (VHR) satellite images especially in complex urban scenes. In this paper, we develop a superpixel algorithm as a modified edge-based version of Simple Linear Iterative Clustering (SLIC), which is here called ESLIC, compatible with VHR satellite images. Then, based on the modified properties of generated superpixels, a heuristic multi-scale approach for building extraction is proposed, based on the stereo satellite imagery along with the corresponding Digital Surface Model (DSM). First, to generate the modified superpixels, an edge-preserving term is applied to retain the main building boundaries and edges. The resulting superpixels are then used to initially refine the stereo-extracted DSM. After shadow and vegetation removal, a rough building mask is obtained from the normalized DSM, which highlights the appropriate regions in the image, to be used as the input of a multi-scale superpixel segmentation of the proper areas to determine the superpixels inside the building. Finally, these building superpixels with different scales are integrated and the output is a unified building mask. We have tested our methods on building samples from a WorldView-2 dataset. The results are promising, and the experiments show that superpixels generated with the proposed ESLIC algorithm are more adherent to the building boundaries, and the resulting building mask retains urban object shape better than those generated with the original SLIC algorithm.

Keywords: superpixels; building extraction; DSM refinement; stereo satellite imagery; multi-scale segmentation; SLIC; ESLIC

1. Introduction

Segmentation, as an integral and basic part of computer vision and image processing, literally means to segment the image into meaningful pieces or into the "ingredients" that a human mind unconsciously does. Numerous studies have been published on the concept of image segmentation but still it is a challenging problem to naturally segment the objects in an image. Regarding the implementation of segmentation algorithms, different types of paradigm have been developed [1,2]. The early reported algorithms comprised region-based, edge-based and thresholding methods [3], mainly employing image pixels as the basic component of the image. Therefore, generally the traditional segmentation algorithms are known as pixel-based approach [4]. Pixels are not natural entities and not efficient representative of the captured space. Since early in 2000s, Blaschke et al. discussed the question: 'what is wrong with pixels?' [5]. Then, they introduced a new paradigm in image segmentation known as Object Based Image Analysis (OBIA) which led to the development of object-oriented GIS and remote sensing software. At the same time, recently we have been facing a rapid progress in Very High-Resolution (VHR) image analysis and a high availability of data which contain higher spatial and spectral information than before. Simultaneously, new image processing techniques started meeting the need for faster and more powerful algorithms and to deal with the huge amount of spatial and spectral data. Consequently, a set of methods has been developed called superpixels [6]. Superpixels are small patches of pixels representing homogenous regions in an image. Derived from a low-level grouping process, they can represent an image in a more abstract way. Additionally, superpixels comprise more meaningful information than that of individual pixel and can be used as an input for further processing algorithms. This leads to an increase in speed, simplicity as well as efficiency of computation. Many of so-called superpixel methods were originally known as interactive methods through which the segmentation starts from an initial predefined segment in the image, followed by a convolutional algorithm [7,8]. Such algorithms such as Random Walk [9] and level set [10] were commonly used in medical image analysis from the early stages. However, during the last years these algorithms have been improved to be more automatic starting from regularly distributed random pixels in the image as superpixel centers which are called 'seeds'. The seeds are then updated in each iteration to form the final superpixels [6]. Accordingly, they are generally known as over-segmentation methods; segmenting an image into small regions which are smaller than objects.

Recently regarding the advantages of superpixel algorithms, many studies have been performed for different remote sensing applications including optical imagery as well as SAR and hyperspectral images with a diverse range of targets [11–13]. For satellite imagery from urban areas particularly building boundaries are critical and have been the main target of many researches in urban remote sensing. The accuracy and reliability of classification or object extraction, resulting from the superpixel-based methods, are highly dependent on the accuracy of the superpixel generation algorithm. Consequently, boundary adherence is an essential factor in the superpixel method selection.

Though here we are focusing on superpixel algorithms, to maintain the comprehensiveness of this review, it is worth mentioning Deep Learning techniques which have received considerable attention over the past few years [14]. Deep learning techniques are basically machine learning methods with multi-layer (deep) convolutional neural networks, that can deal with huge amount of input data to predict the output. They are increasingly being used in segmentation and classification problems in the field of remote sensing applications as well and have been reported to get promising results [15].

In this paper, inspired from the state-of-the-art superpixel generation algorithm, we propose an improved method adapted to VHR multispectral satellite images in urban scenes. Since buildings are of high interest in urban remote sensing, in the proposed algorithm we are particularly focusing on the building boundaries.

This work is composed of two main parts: modified superpixel generation and building mask extraction. In the next section, Section 2, first we have a review on some of the most popular superpixel segmentation algorithms reported in the literature, then a modification on the SLIC superpixel algorithm is proposed. In Section 3, based on the modified properties of the derived superpixels, a multi-scale approach for building mask generation is applied. The results and discussion come in Sections 4 and 5 respectively, and Section 6 conclude the findings.

2. Superpixel Generation

2.1. Background and Motivation

As mentioned above, most of so-called superpixel algorithms are modified versions of popular segmentation methods and have the same basics. In this part we review some of the most widely known superpixels and their concepts.

Considering the image as a graph, is the basic idea behind a set of superpixel generation algorithms in computer vision [16–19]. In 2000, Shi and Malik [18] developed a segmentation algorithm based on graph theory known as Normalized Cuts (Ncuts). They consider the image segmentation as a graph partitioning problem based on a normalized cut criterion which measures both the total dissimilarity between the different superpixels as well as the total similarity within the groups [18]. Later some studies have been proposed superpixel algorithms based on the multi-scale formation of Ncuts in [20,21], and more recently [22] in which a Linear Spectral Clustering (LSC) superpixel algorithm as a normalized formation of Ncuts is proposed that measures the color similarity and space proximity between image pixels. However, generally the Ncuts algorithm, with the complexity of $O(N^2)$, is computationally expensive especially when the number of pixels increases.

Graph cuts superpixel segmentation, first proposed by Boykov et al. in 2001, is one of the common superpixel segmentation methods based on graph and energy function optimization [17]. Given a set of pixels as graph nodes, a cut is a subset of edges and is equivalent to the minimum cost and it happens at the image boundaries. The algorithm has the complexity of $O(N^2)$ when N is the number of pixels, therefore it is not a fast algorithm [23]. Various superpixel algorithms have been reported in the literature based on the same idea [16].

Another popular segmentation method with the concept of modeling an image as a graph, is the Random Walk (RW) algorithm proposed by Grady [9]. The basic idea of the RW method is to model the image as a graph in which each pixel corresponds to a node which is connected to neighboring pixels by edges, and the edges are weighted to reflect the similarity between the pixels [9]. The RW was originally developed as an interactive image segmentation using the foreground and background seeds by the user. As a modification of the RW method, Shen et al. in [24] proposed an algorithm called Lazy Random Walk (LRW). In the RW the model ignores the whole relation of the current pixel to the other seeds, therefore it may fail in weak boundaries and complex textures. To make it globally in comparison to the RW which operates locally, they add a 'self-loop' over the graph vertex and called the algorithm 'lazy'. The complexity of LRW algorithm is $O(N^2)$ which is relatively high. However, as an advantage they reported this algorithm to be boundary adherent even in weak boundaries and complex textures [24].

The second group of superpixels are those from the family of active contour and level set methods. Levinshtein et al. in 2009 [25] proposed the Turbopixel superpixel algorithm based on a restricted form of level set [10]. Level set segmentation as a very common algorithm in medical image processing has been increasingly applied to image segmentation in the past decade [18,20,26]. The idea is to represent the boundary of a superpixel in an image, as a curve C with normal vector N and the speed of curve evolution which is then $\frac{\partial C}{\partial t} = SN$. Considering the signed Euclidean distance of each image pixel to the closest point on the boundary, a pixel's distance is positive if the pixel is inside the region and negative if not. Accordingly, it is zero for the pixels on the region boundaries. Therefore, through the level set algorithm the curve is iteratively evolved and finally the zero-level curve is regarded as the region boundary [25]. Through the Turbopixel algorithm, starting from the uniformly placed initial seeds, a localized level set is applied. The algorithm complexity has been reported to be roughly linear to the total number of pixels; $O(N)$.

Generally, four basic principles are considered to establish a superpixel algorithm; connectivity, uniform size and shapes, edge preserving and getting non-overlapped superpixels [25]. Besides, different superpixel algorithms are commonly compared and evaluated regarding three main principles: boundary adherence, compactness, computational complexity, or speed [23,27].

In 2012 Achanta et al. [27] developed the Simple Linear Iterative Clustering algorithm (SLIC), as localized k-means. The idea is to locally cluster the image pixels regarding the computed distance using both spatial and color proximity to the initial centers. SLIC has shown to be a fast and powerful algorithm with the highest boundary adherence and the lowest complexity [23]. Accordingly, SLIC has been recently of great interest to researchers in computer vision and remote sensing [11,13,22,28].

There are other segmentation algorithms under the category of superpixel generation methods such as Lattice [29], SEEDs [30], Quick-shift [31] as a kernelized mean-shift algorithm [32], which in comparison to the above mentioned algorithms have relatively poor performance and/or higher complexity and lower speed [23]. Besides, in the literature one may find some other algorithms, but they are mainly variations or modifications of the cited major methods.

In the following part we propose a modified superpixel algorithm for VHR satellite images. The main target of our study is building detection, therefore here we focus on urban scene segmentation.

2.2. SLIC

Basically, SLIC is a localized iterative k-means algorithm [27]. The algorithm starts from a regular grid of seeds. The number and size of superpixels depend on the number of initial seeds. For each seed point a neighborhood with size 2 × S is defined, where S is the initial grid interval, and the nearest neighboring pixels are labeled to be in the same superpixel. The distance measure, D in Equation (1), is composed of both color distance d_c in CIELab space, as in Equation (2), and spatial Euclidean distance d_s, as in Equation (3), so the effect of each parameter can be controlled by the weighting value m:

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

$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \tag{2}$$

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

where $[l, a, b]$ are color coordinates in the CIELab color space and $[x, y]$ are pixel's position in the image. According to the standard SLIC algorithm, at the beginning each pixel in the image is likely to belong to 4 regions (except for the peripheral seeds), Figure 1.

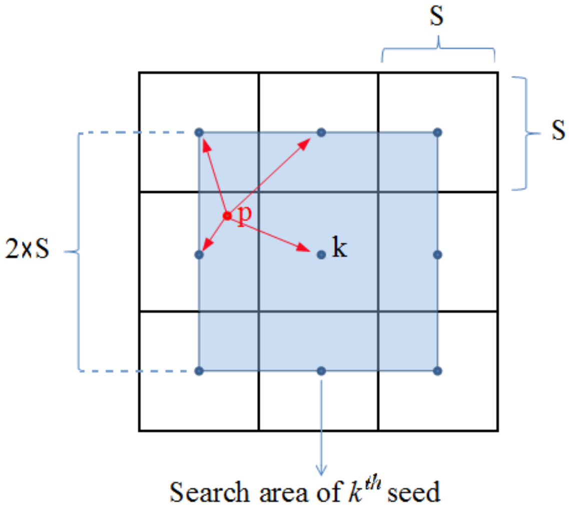

Figure 1. Each seed k has a search area of 2S × 2S to generate superpixels with size S × S, and each pixel p is likely to belong to 4 seeds.

Starting from the first seed, the nearest neighboring pixels with a distance lower than a very large initial predefined value is assigned to the seed and in the meanwhile the initial distance value for those pixels is updated. After formation of the first superpixel which is obviously expected to have a regular shape, the second seed point is inspected in its own neighborhood and the distances as well as the labels for the pixels through the common area with the previous neighborhood are updated again. The procedure continues until every pixel in the image belongs to its nearest and at the same time the most spectrally similar superpixel. This is iteratively applied and after each run, the seed centers are updated. Through this procedure, m as a weight parameter, tunes the relative importance of each term in Equation (1) and has a significant rule on the shape of final superpixels. When m is small, the resulting superpixels have higher boundary adherence but shape and size regularity decreases. SLIC is $O(N)$ complex. In comparison to the other superpixel segmentation algorithms, SLIC shows to have higher speed, more boundary adherence, and a relatively low under-segmentation error [23].

2.3. Problem Statement and Modification

Figure 2 shows the resulting SLIC superpixels for different parameters of m and N on an urban scene in central Munich. It can be seen that for large values of m, superpixels are regular-shaped and do not fit especially to the weak object boundaries. On the other hand, smaller values of m make the algorithm too sensitive to the color changes and again the superpixel boundaries do not fit properly to the real meaningful objects. Accordingly, SLIC tends to fail in the building boundary reconstruction in urban scenes from satellite images due to various reasons such as complexity of city structures and intricate roof buildings, shadows, and sometimes abrupt illumination changes. In the standard form of SLIC, the shape of final superpixels is highly affected by the color similarity of different surfaces on roof tops and surrounding ground in the CIELAB color space [27]. Thus, considering the availability of valuable spectral information, for instance 8 spectral bands of WorldView-2 images, it is reasonable to modify the spectral feature space in the algorithm. In this paper, we propose a modified superpixel algorithm, from now on called Edge-based SLIC and noted as ESLIC, to compensate for the drawbacks of applying standard SLIC on VHR satellite images in urban areas.

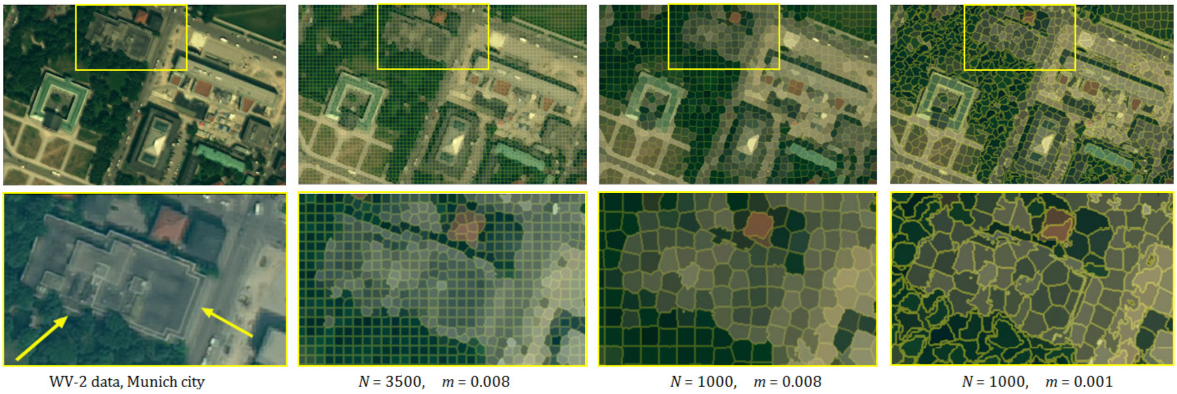
WV-2 data, Munich city
$N = 3500, \quad m = 0.008$
$N = 1000, \quad m = 0.008$
$N = 1000, \quad m = 0.001$

Figure 2. Results of SLIC superpixels on a sample WV-2 urban scene. The algorithm tends to fail on weak boundaries which are pointed by arrows.

2.4. ESLIC

As discussed above, to increase the capability of the superpixel algorithm to distinguish different surfaces with similar optical behavior, in the proposed algorithm, three modification are introduced. First, we use the original spectral bands instead of the CIELAB colors. In this way, the input feature vector to calculate color distance, from a 3-dimensional vector is turned to a higher dimensional one consisting of the 8 spectral bands for WorldView-2 data.

$$d_c^2 = \sum_{i=1}^{i=n} \left[(g_i - g_{ki})^2 \right] \tag{4}$$

where g_i refers to the intensity value of a pixel in the ith band and k is the index for the kth seed point.

Besides, urban areas having complex textures may cause significant intensity fluctuations which can lead to weak boundaries. This makes it more difficult for the algorithm to meet the boundary adherence condition in the satellite images. Therefore, as the second modification, to get sharp edges with the most coincidence to the real object boundaries, an edge indicator is introduced to the superpixel shape formation phase in the proposed modified algorithm. Therefore, not only the spectral variations through the image, but also the geometry of study area are considered. To allow this indicator to affect the shape of superpixels, a new term related to the edge indicator, E, is added as an extra element in the input feature vector. This leads to an increase in dimensionality of feature space. Therefore Equation (2) will change into the form:

$$d_c^2 = \sum_{i=1}^{i=n} \left[(g_i - g_{ki})^2 + E_i \right] \tag{5}$$

An edge indicator E for each pixel in the image generally can be defined for example using image gradients, image Laplacian or other expressions like in [33]. In this study, we have first tested the magnitude of gradient and Laplacian as edge indicator in Equation (5). However, the Laplacian image is too sensitive to noise as well as intensity changes and the resulting superpixels are rough and noisy in both shape and size, as shown in Figure 3. Image gradient represents directional change in the intensity or color in the image. Therefore, on weak boundaries which is a common case in satellite urban scenes, it will have a very small magnitude. That will reduce the power of the algorithm for boundary adherence and can lead to a leakage of superpixels near the boundaries. As we need a powerful constraint on the building boundary for superpixel formation, and according to the fact that SLIC is basically a region-based clustering algorithm, we reinforce the edge indicator with a textural term in Equation (5).

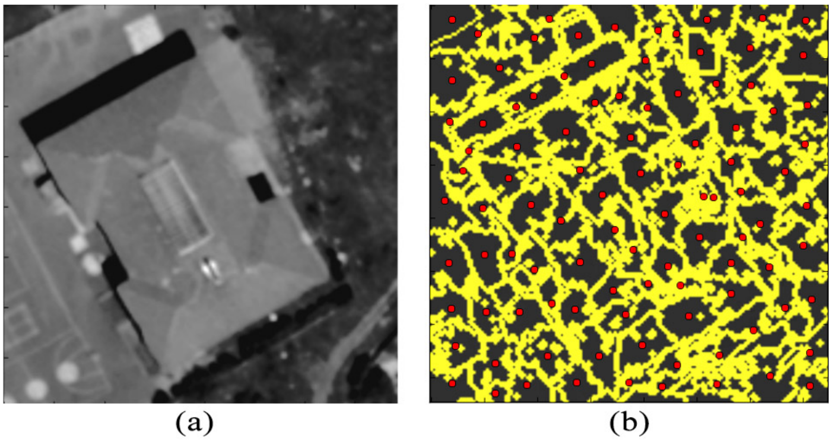

Figure 3. A building in a PAN image (**a**), and corresponding superpixels generated using a Laplacian constraint (**b**). Red dots are superpixel centers. Laplacian constraint in the formation of superpixels makes the algorithm too sensitive to noise and color fluctuations.

Texture is a description of the image homogeneity based on local spatial variations of intensity or color brightness that has an important role in remote sensing [34]. Especially the gray level co-occurrence matrix (GLCM) is a classic common method which calculates the local correlation of pixels to obtain the texture feature values. Among all reported feature statistics based on GLCM to describe the image texture features [35], we found contrast a proper reinforcement for the edge-term in our algorithm. Texture contrast (CON) which can be described as 'generalized gradient' [36], for a 2D image $f(i,j)$ is defined by Equation (6):

$$CON = \sum_{i}^{N} \sum_{j}^{N} (i-j)^2 \, P(i,j) \tag{6}$$

where i, j refer to the image coordinates and P refers to the GLCM, i.e., square matrix whose size represents the probability of the gray value g_1, distanced from a fixed spatial location relationship (size and direction) to another gray value g_2 [34], and can be expressed as:

$$P(i,j) = \frac{\#\{[(i_1, j_1), (i_2, j_2)] \in S | f(i_1, j_1) = g_1 \& f(i_2, j_2) = g_2\}}{\#S} \tag{7}$$

where S is the set of pixels with a certain spatial relation in the region. As a result, the edge indicator term E in Equation (5) can be defined by:

$$E = d_G^2 + d_{GLCM}^2 \tag{8}$$

where d_G and d_{GLCM} represent the difference of image gradients and the difference of texture GLCM for each pixel, respectively.

As the third modification, the connectivity constraint is altered. Connectivity constraint is usually one of the main factors considered in superpixel generation procedures. Many existing superpixel methods, such as Normalized Cuts, Graph Cuts and Turbopixels, meet this constraint inherently through the algorithm [17,18,25]. SLIC in the standard form doesn‘t explicitly enforce connectivity and uses a connected components algorithm as a post-processing stage before final superpixel generation [27]. This causes some details to be ignored or some unwanted bulges and artifacts on the boundaries to appear. Here in contrast, we suggest enforcing the decomposition before final superpixel generation. This is well adapted to preserve fine details in satellite images from very complex city scenes. It is also able to improve the boundary representation on which further experiments are based. In Figure 4, the effect of using the proposed ESLIC, is shown.

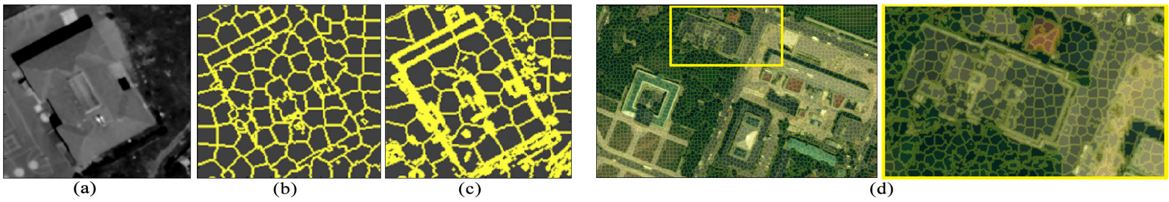

Figure 4. Effect of gradient term on the shape of superpixels; building in a Pan WV-2 image (**a**), superpixels resulting from SLIC (**b**), superpixels resulting from ESLIC (**c**,**d**). The results shown in (**c**,**d**) can be compared to Figures 2 and 3b respectively.

In this research we have used our modified superpixel algorithm in a heuristic approach for building extraction from VHR satellite images.

3. Building Mask Extraction

Building mask generation is a challenging problem in urban remote sensing. Due to the importance of buildings as the main objects in urban areas, various algorithms on building detection and extraction have been developed and many researches have been reported in the literature since the 1980s [37,38]. Recent advances in computational information technology as well as in development of high-resolution sensors lead to a wide range of heuristic and innovative approaches for urban object detection in satellite images. We can divide all the reported algorithms into two general categories; data-driven and model-driven methods. Although many of reported algorithms are hybrid models of data-driven and model-driven schemes. In the model-driven scenario, first some predefined models are constructed for the buildings and then they are projected onto the image to be adapted to the extracted features [39]. When the spatial resolution of the image is relatively low in comparison to the object extent, the model-driven approaches have better performance [40]. However, data-driven methods are usually based on image grouping or segmentation. Therefore, compared to the model-driven methods, they are more adaptive to complex building shapes and types. In this paper, we propose a data-driven method for building extraction, supported by our modified superpixel segmentation, ESLIC.

3.1. Methodology

The heuristic method we developed in this research is based on VHR satellite images and relies on the proposed superpixel algorithm described above. As shown in Figure 5 the workflow starts with the stereo images as input data. Then, using the Normalized Difference Vegetation Index (NDVI) and shadow extraction from Pan-Sharpened image (PS) an initial coarse mask is generated from the normalized DSM (nDSM). This initial mask is then imposed on the PS and ESLIC is applied on building candidate areas to generate superpixels with different sizes. Finally, multi-scale superpixels are integrated to generate the final building mask. In the next sub-sections, the workflow is described in detail.

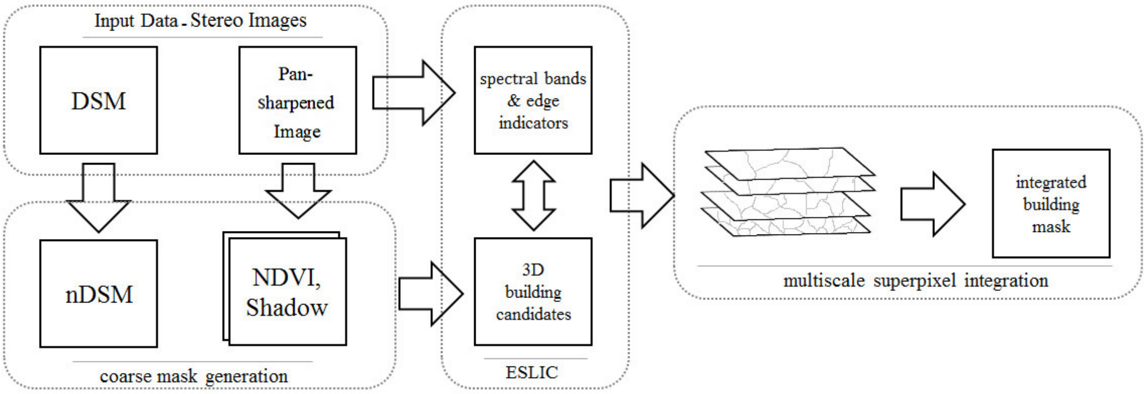

Figure 5. Workflow of building mask extraction supported by multi-scale ESLIC.

3.1.1. Data Preparation

Digital surface models (DSM) can be efficiently generated from VHR optical stereo images with the Semi-Global Matching (SGM) algorithm first proposed in [41], and later improved for satellite data in [42]. The DSM is used here for two main purposes; to generate the orthorectified image to be segmented, and to extract building candidates. As described earlier in Section 2.4, ESLIC employs spectral as well as spatial and geometric features derived from input images. Therefore, first the input satellite image needs to be PS and orthorectified. Then, shadows and vegetation (NDVI) are removed from the nDSM to estimate building locations.

3.1.2. DSM Refinement

Since stereo DSMs, are suffering from various local defects through errors in stereo matching and 3D model generation, and due to different natural phenomena such as illumination variations and occlusions, therefore containing holes so-called DSM voids; in the first step the DSM needs to be refined to get a seamless model. To fill the voids a range of algorithms have been proposed, mainly including interpolation alone or along with using an auxiliary DSM. Here as in [43] we employ a segmentation-based filling approach, shown in Figure 6.

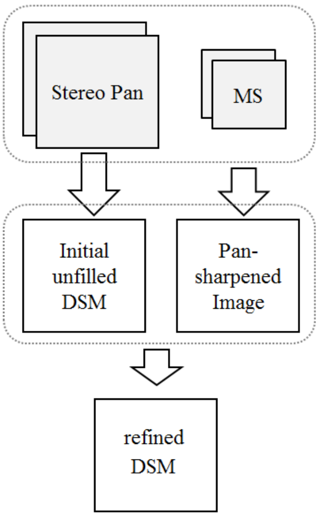

Figure 6. Data preparation; DSM refinement through void filling using spectral information from multispectral image.

In this way, no external DSM is needed to fill the voids. First a neighborhood around every void in the unfilled DSM and its corresponding area in the multispectral image is defined. Then, these areas in the image are segmented using the ESLIC algorithm, so that according to the corresponding superpixel, a label is assigned to each cell in the voids. Finally, these void cells in the DSM are filled by the interpolated value of their co-labeled neighbors.

3.1.3. Multi-Scale Building Extraction

Having a coarse building mask as described in Section 3.1.1, regarding the maximum elevation in the study area, a 2D initial mask is created. The mask is overlaid on the image and the modified superpixel algorithm is applied on the candidate areas to reduce the computation load and to increase the focus on the specified regions. To get rid of the parameter dependency of the superpixel shape and size, a multi-scale approach is used and ESLIC is applied on every scale. That gives multiple sets of superpixels with different size parameters. Our proposed ESLIC algorithm guarantees that in each scale there are some fixed small patches, say anchors, in common with the other scales, containing edge points.

As mentioned in Section 2.4, we remove the connectivity constraint in ESLIC. This removal benefits our approach in two ways. First is to highlight and distinct the superpixels which are completely inside the mask. As a result, of this constraint, surrounding superpixels with high probability of comprising boundaries as well as leaked ones, are removed and interior superpixels are regarded to be labeled as building parts. In addition, using the edge-term in the proposed algorithm through Equation (8), make the superpixels follow the basic shape of the building skeleton say ridge lines and its boundaries. This reduces the risk of uncertain superpixels, i.e., those that are not completely inside the building area, and saves only the safe ones as small pieces of buildings. These remaining small pieces obtained from every scale, are then integrated larger area, here named as “object core” is created, such as the sample shown in Figure 7.

Next, we generate a mean-image, so that every pixel in this image receives the average value of the corresponding intensities regarding the new integrated superpixels. Expectantly, the object core in the new segmented image gets homogeneous intensities and textures. Therefore, if this mean-image is again segmented using the ESLIC, as a result, the edge constraints on the interior parts are automatically removed and only surrounding superpixels follow the building boundary shape. Since the superpixels especially those containing weak boundaries or high textured areas, are composed of several smaller parts, first we impose detachment. This is to remove exterior small patches regarding the shape of object core. The main key to the success of the algorithm in this stage is the reinforced edge-term, in Equation (8), which provides a continuous and seamless constraint along the building boundaries.

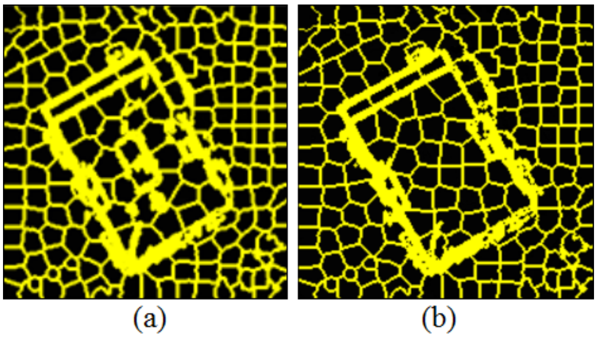

Figure 7. ESLIC on the pan-sharpened image (**a**), ESLIC on the mean-image generated using the integrated superpixels on the object core (**b**).

3.2. Remark

In Figure 8 a sample building in the Munich dataset is shown and one can see the confusion problem through some boundaries and building edges. Due to the geometry of perspective projection, the shape and position of objects in the optical satellite images are exposed to relief displacement or height distortion. Displacement increases with the radial distance from nadir point and with the height of objects. Orthorectification transforms the central projection of the image into an orthogonal view of the ground, thereby removing the errors and distortions including relief displacement. However usually relief distortion cannot fully be corrected in the orthorectified image, because of deficiencies in the DSM, registration, etc. Therefore, to get rid of this problem, we can use the original images instead of the orthorectified one, as the input in the algorithm (Figure 5). If we use un-orthorectified image as input, we need a co-registered height model and DSM cannot be employed anymore. Therefore, instead we employ the height map (HM), which is the first product of SGM following epipolar lines using sensor parameters or Rational Polynomial Coefficients (RPC). This epipolar rectification through the height map generation, makes it co-registered to the master image. Like the stereo DSM, this HM contains some gap areas and voids, so the filling and refinement as described in Section 3.1.2, is still essential. The main advantage of this replacement is that the edges and boundaries in the un-orthorectified image are sharper; so the artifacts around the urban objects are eliminated and the ESLIC algorithm may preserve building boundaries more reliably. For each HM, in the same way described earlier, after removing NDVI and shadow from corresponding master image, the regions as building candidates are extracted first. Then, the master image is segmented, and the corresponding HM is filled. Both are the input for ESLIC superpixel segmentation.

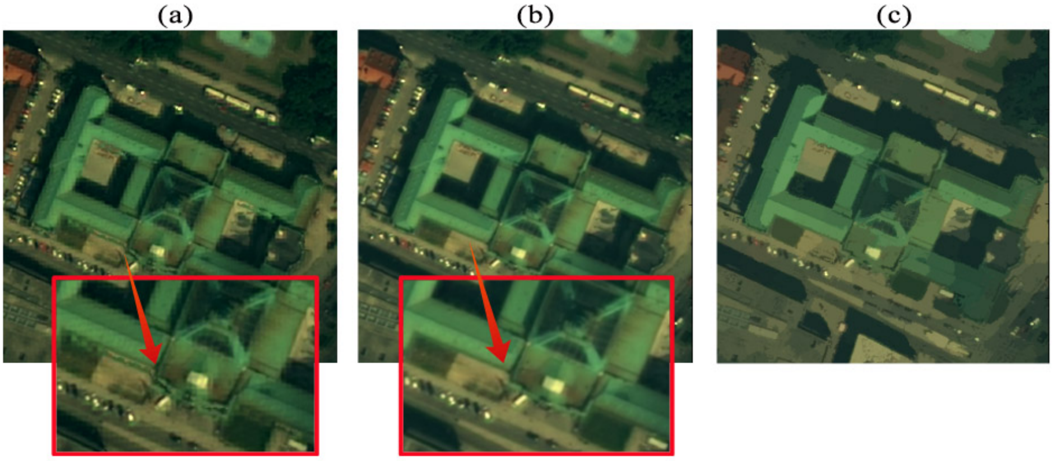

Figure 8. Artifacts around building boundaries in the orthorectified image (**a**), made us use the un-orthorectified image (**b**). (**c**) shows the result of ESLIC superpixel segmentation on the image (**b**).

4. Results

4.1. Data

The input data used in this study are satellite images from Munich and Istanbul city taken with WorldView-2 in 2012. The DSMs were generated from panchromatic stereo images with 0.50 m resolution. To perform the experiments, we picked different buildings in the cities from the DSM and the corresponding parts in panchromatic and multispectral images.

4.2. Experimental Setup and Results

To generate the modified superpixels with our proposed algorithm, only two parameters should be set as in the original SLIC parameters; number of superpixels (N) and the weight parameter (m). As described in Section 3, we apply the algorithm in multiple scale, therefore setting a precise value for N no longer matters critically. Instead, the algorithm starts with an initial N and continues for larger values. Obviously, increasing N creates smaller superpixels and the maximum value of N would be equal to the image size, resulting in superpixels each of which is equivalent to a single pixel in size.

Regarding Equations (5) and (8), 8 spectral bands of WV-2 images are used as well as an edge-term containing the gradient values and texture features, GLCM, for each pixel to generate the ESLIC superpixels. Since GLCM has high computational complexity and makes the algorithm rather slow, to reduce computational burden, we apply the ESLIC algorithm only on the selected area containing building candidates. The candidate areas are extracted initially from the nDSM, after shadow and vegetation (NDVI) removal. For shadow detection from multispectral satellite images, many algorithms have been proposed. We used the spectral index in [44] to automatically detect shadows using a ratio defined by the Blue band, B, and two Near-Infrared bands, NIR1 and NIR2, ((NIR2-Blue)/(NIR2-Blue)-NIR1). Proper values for N are then set regarding the approximate size of buildings and is determined by the number of pixels, n_p, divided by the initial minimum desired size of superpixels, s; ($\frac{n_p}{s}$). The latter value depends on the scene structure as well as the Ground Sampling Distance (GSD). Accordingly, in the WV-2 data with 50-centimeter GSD, the size to be regarded as a meaningful building part is considered to be around 16 meters equivalent to about 80 pixels. On the other hand obviously the minimum size of superpixels should be smaller than the smallest building or building part in the scene.

We apply this for 5 scales and the accumulation of these labeled parts is used for the building mask extraction. Multi-scale ESLIC-based building extraction is illustrated in Figure 9. After removing external and border superpixels regarding the initial mask, only internal superpixels are kept for each scale. These parts are then integrated (Figure 9f) and final building superpixels are extracted from the mean-image in the same way. Finally, using morphological operators the remaining tiny holes inside the mask are covered and compensated. The results for the test dataset is shown in Figure 10.

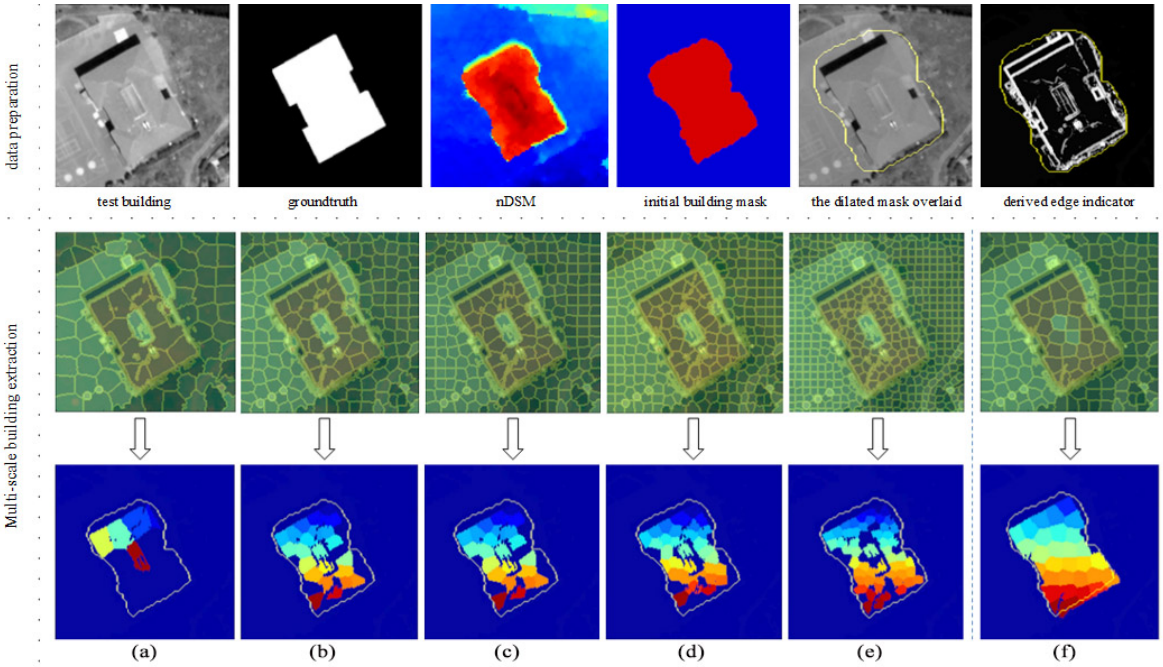

Figure 9. The first row shows the data preparation phase results for a test building. In the second row the mean-images generated from multi-scale ESLIC superpixels are shown. The last row illustrates how the most probable building parts are derived and kept from the initial rough masks; starting from bigger scale and larger superpixels to the smaller one (**a**–**e**), and how the integrated superpixel is overlaid on the mean-images on each scale (**f**). In the last row each color represents a superpixel.

4.3. Quantitative Evaluation and Comparison

Generally, in computer vision, computational vision algorithms are evaluated in two different ways. In the first approach, the algorithm is evaluated in the context of a particular task [45]. It means to measure the contribution of the algorithm to the higher-level procedures, for example in remote sensing, the image segmentation algorithm should be evaluated regarding its contribution to the quality of the final detected objects. In the second approach, the performance of an algorithm is evaluated regarding a given ground truth.

Using ground truth or generally reference data for superpixel segmentation evaluation, commonly three measures are used: Boundary Recall (BR) as in Equation (9), measuring the fraction of ground truth boundaries recovered by the segmentation results;

$$BR(S,G) = \frac{\sum_{p \in dG} \left(min_{q \in dS} \, ||p-q|| \leq \epsilon \right)}{|dG|} \tag{9}$$

Under-Segmentation Error (UE), measuring the percentage of pixel leakage from the ground truth boundaries as in Equation (10);

$$UE(S,G) = \frac{\sum_i \sum_{k: s_k \cap g_i \neq 0} |s_k - g_i|}{\sum_i |s_i|} \tag{10}$$

and the highest Achievable Segmentation Accuracy (ASA), as the fraction of labeled pixels that are not leaked from the ground truth boundaries, defined in Equation (11) [46];

$$ASA(S,G) = \frac{\sum_k max_i\, |s_k \cap g_i|}{\sum_i |g_i|} \quad (11)$$

where $S = s_1, s_2, s_3, \ldots, s_{n_s}$ refers to the resulting segments and $G = g_1, g_2, g_3, \ldots, g_{n_g}$ presents the ground truth. The boundaries in the segmentation result and in ground truth are represented by dS and dG respectively. Table 1 shows the quantitative results for the test buildings presented in Figure 10.

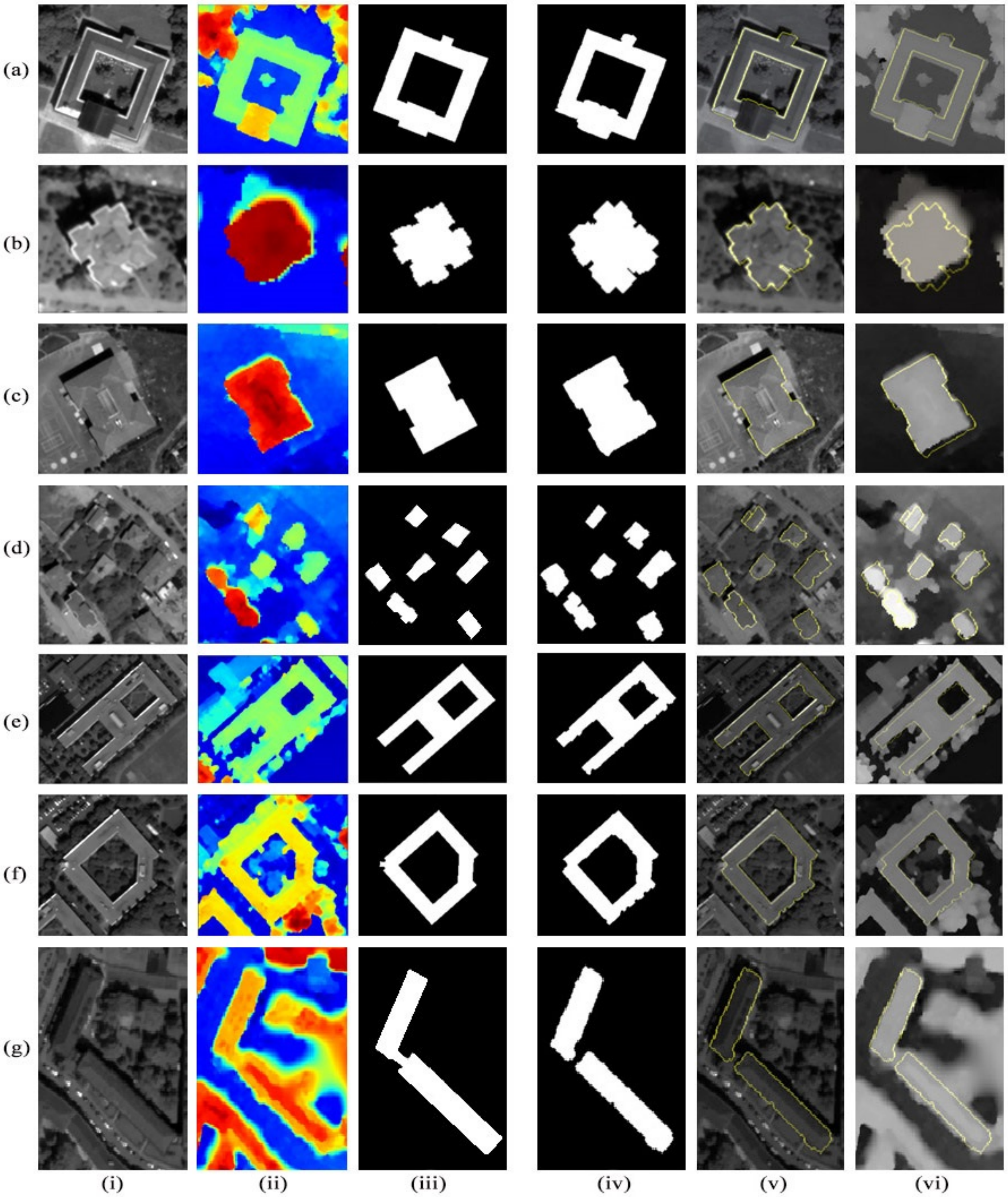

Figure 10. Results for 7 test data (**a**–**g**); (**i**) pan image, (**ii**) DSM, (**iii**) ground truth, (**iv**) result as extracted mask, (**v**) result as footprint overlaid on the pan image, (**vi**) overlaid on the DSM.

Table 1. Evaluation of the generated building masks using ESLIC and comparison to the SLIC and level set results, for the building samples a–g in Figure 10.

Test Data	ESLIC			SLIC			Level Set		
	UE	BR	ASA	UE	BR	ASA	UE	BR	ASA
a	0.053	0.83	0.94	0.070	0.85	0.92	0.065	0.89	0.92
b	0.052	0.93	0.96	0.068	0.93	0.89	0.060	0.90	0.86
c	0.048	0.98	0.95	0.065	0.98	0.91	0.059	0.91	0.90
d	0.058	0.88	0.93	0.070	0.90	0.85	0.068	0.85	0.89
e	0.032	0.93	0.98	0.064	0.94	0.93	0.054	0.92	0.94
f	0.042	0.94	0.95	0.060	0.93	0.94	0.061	0.93	0.95
g	0.038	0.95	0.95	0.058	0.92	0.92	0.070	0.93	0.91
mean	0.046	0.92	0.95	0.065	0.92	0.91	0.062	0.90	0.91
STD	0.009	0.050	0.016	0.005	0.040	0.030	0.005	0.028	0.030

In the last three columns of the table the quantitative evaluation result of the obtained building masks from another primitive superpixel algorithm, level set, is shown. As described earlier in Section 2.1, level set is originally introduced in the context of active contour model. It has achieved good performance in image segmentation and boundary detection in computer vision and has been recently tested on VHR satellite image in [47].

5. Discussion

Table 1 shows that ESLIC generally achieves satisfying results for the test dataset used in this research. It demonstrates that in comparison to SLIC and level set algorithms, ESLIC superpixels have lower UE, meaning lower percentage of pixel leakage from the ground truth boundaries. This advantage is due to using textural as well as spectral features in the ESLIC superpixel generation. In the same way, comparing the ASA and BR measures in the table, approves the superiority of ESLIC algorithm to the other methods. It means that ESLIC superpixels are generally more adherent to the building boundaries and less leaked from the ground truth boundaries.

In Figure 11, one can also visually compare the resulting masks from ESLIC, SLIC and level set for the 7 test buildings. It is clear that overall performance of ESLIC is higher than the other two tested methods. In addition, inspecting the visualization of generated superpixels, in comparison to the SLIC, ESLIC has the advantage of using a powerful edge constraint. This particularly makes sense through the multi-scale approach. Scaling typically leads to some smoothing and approximation, therefore the SLIC which is solely based on spectral features may generate quite different superpixel shapes in each scale (Figure 12). However, the strong point of our proposed algorithm is the fact that regardless of the desired scale, building boundaries as well as main ridge lines are preserved by the superpixel shapes. It increases the reliability and robustness of the ESLIC algorithm for building mask generation.

On the other hand, comparing level set method to ESLIC and SLIC algorithms, it has more parameters to be set, and many parameters must be tuned depending on the dataset. While in the proposed multi-scale algorithm, only a single weight parameter (m), should be set for ESLIC and SLIC segmentation.

Since our proposed algorithm has data-driven approach, the results can be affected by occlusions in complex situations such as a building being partly covered by trees. A sample is shown in Figure 10f in which the lower right side of the building is slightly covered by surrounded single trees. The indentation along the extracted building boundary in this side can be seen clearly in Figure 10f-iv. These kinds of occlusions can be partly corrected by more intelligent and context-aware NDVI and shadow removal methods. However, some will remain the challenging issues in the field of data-driven object detection approaches.

Here we tested the proposed method for WorldView-2 stereo images; however, the ESLIC algorithm is compatible with every multispectral dataset. Besides, the proposed algorithm can be applied on the average size dataset as well. However, when dataset increases in size, the algorithm runs slower, because the complexity of texture computation increases. To avoid this problem, we applied the algorithm only on the cropped area containing building candidates.

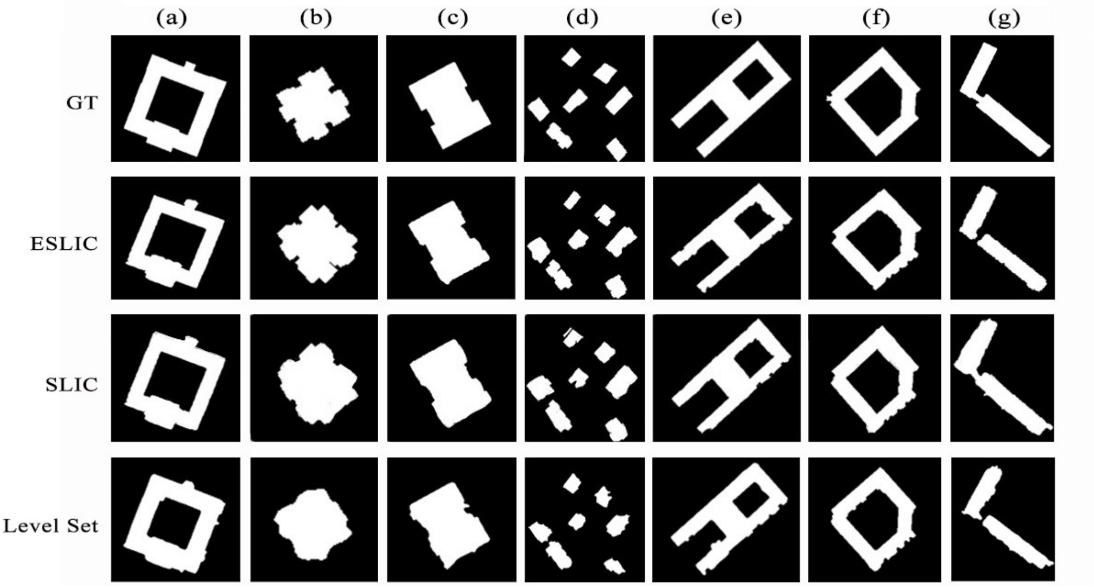

Figure 11. Results of ESLIC in comparison to the ground truth (GT), SLIC and level set results for 7 dataset (a–g) tested in Table 1.

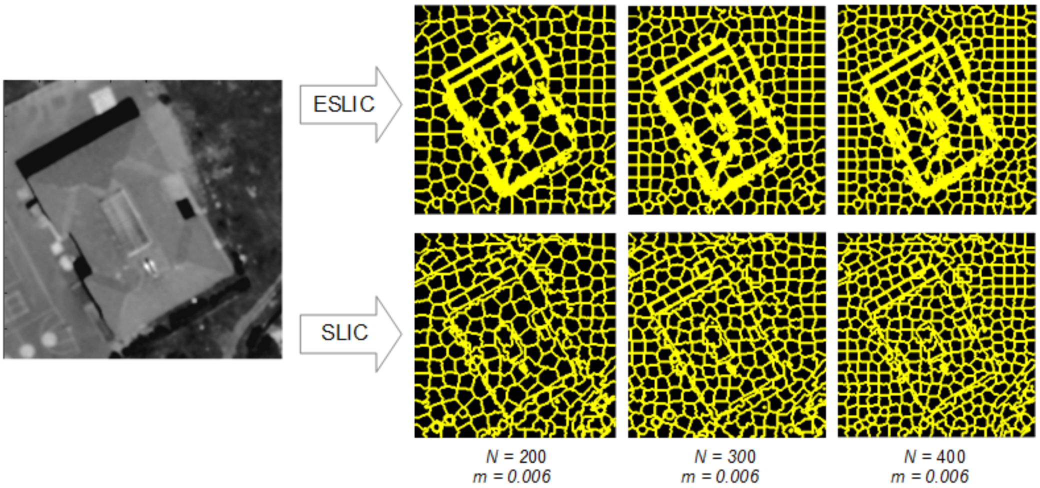

Figure 12. ESLIC results (**top**), vs. SLIC superpixels (**bottom**), both for the same scales. ESLIC can preserve boundaries and main ridge lines regardless of the scale.

6. Conclusions

In this paper, we proposed a modified superpixel algorithm, adapted to VHR remote sensing images. To preserve weak boundaries in the image and to increase the boundary adherence of superpixels, an Edge-based SLIC, named ESLIC, has been developed. The algorithm shows to outperform the standard SLIC algorithm for VHR multispectral satellite images, due to the extra spectral and geometric features introduced to distance computation. Then, a heuristic multi-scale building detection method based on the modified superpixels is proposed, which employs the approximate location from the DSM and the spectral and textural features derived from the corresponding image. The proposed ESLIC is also used for DSM refinement as a preprocessing phase to fill the unwanted gaps and void areas caused by the occlusions, shadows, etc.

In this research the algorithm has been first tested on orthorectified image. Therefore, the artifacts due to the relief displacement around the urban objects in the area, particularly the higher buildings, destroyed the boundaries and disturbed the superpixel segmentation result. To solve this problem and create sharper and more precise boundaries, we applied ESLIC on the original image before orthorectification. The results are promising, and this scenario can be also used for an innovative refined DSM generation procedure.

As a conclusion, attaining an improvement in the segmentation of VHR satellite images consequently leads to two main achievements; first to more reliable and accurate results for object/building detection and classification, and second, it can be used for an improvement in the DSM refinement algorithm, i.e., in the void filling procedure. Both achievements are also directly involved in many other applications such as city modeling and 3D object change detection [43,48,49].

Author Contributions: Z.G. was responsible for the main part of the experiment design, coding and writing the paper. J.T. and P.R. supervised the whole process and provided detailed advice during the writing process and improved the manuscript.

Funding: The research was funded by the Space Directorate of the German Aerospace Centre (DLR). Z.G is supported by a DLR-DAAD Research fellowship (No. 57144088).

Acknowledgments: The authors would like to thank Anahid Basiri for her technical support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dey, V.; Zhang, Y.; Zhong, M. A review on image segmentation techniques with remote sensing perspective. In Proceedings of the Technical Commission VII Symposium, Vienna, Austria, 5–7 July 2010.
2. Vantaram, S.R.; Saber, E. Survey of contemporary trends in color image segmentation. *J. Electron. Imaging* **2012**, *21*, 040901. [CrossRef]
3. Pal, N.R.; Pal, S.K. A review on image segmentation techniques. *Pattern Recognit.* **1993**, *26*, 1277–1294. [CrossRef]
4. Blaschke, T.; Lang, S.; Lorup, E.; Strobl, J.; Zeil, P. Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications. *Environ. Inf. Plan. Polit. Public* **2000**, *2*, 555–570.
5. Blaschke, T.; Strobl, J. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GeoBIT/GIS* **2001**, *6*, 12–17.
6. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. *Slic Superpixels*; Technical Report; EPFL: Lausanne, Switzerland, 2010.
7. Li, C.; Kao, C.Y.; Gore, J.C.; Ding, Z. Minimization of region-scalable fitting energy for image segmentation. *IEEE Trans. Image Process.* **2008**, *17*, 1940. [PubMed]
8. Li, C.; Huang, R.; Ding, Z.; Gatenby, J.C.; Metaxas, D.N.; Gore, J.C. A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. *IEEE Trans. Image Process.* **2011**, *20*, 2007–2016. [PubMed]
9. Grady, L. Random walks for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1768–1783. [CrossRef] [PubMed]
10. Cremers, D.; Rousson, M.; Deriche, R. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *Int. J. Comput. Vis.* **2007**, *72*, 195–215. [CrossRef]
11. Xiang, D.; Tang, T.; Zhao, L.; Su, Y. Superpixel generating algorithm based on pixel intensity and location similarity for SAR image classification. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1414–1418. [CrossRef]
12. Zhou, Y.; Wang, H.; Xu, F.; Jin, Y.Q. Polarimetric SAR image classification using deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1935–1939. [CrossRef]
13. Zhang, S.; Li, S.; Fu, W.; Fang, L. Multiscale superpixel-based sparse representation for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 139. [CrossRef]
14. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef] [PubMed]

15. Bittner, K.; Adam, F.; Cui, S.; Körner, M.; Reinartz, P. Building Footprint Extraction From VHR Remote Sensing Images Combined with Normalized DSMs Using Fused Fully Convolutional Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2615–2629. [CrossRef]
16. Felzenszwalb, P.F.; Huttenlocher, D.P. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **2004**, *59*, 167–181. [CrossRef]
17. Boykov, Y.; Funka-Lea, G. Graph cuts and efficient ND image segmentation. *Int. J. Comput. Vis.* **2006**, *70*, 109–131. [CrossRef]
18. Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905.
19. Yi, F.; Moon, I. Image segmentation: A survey of graph-cut methods. In Proceedings of the 2012 IEEE International Conference on Systems and Informatics (ICSAI), Yantai, China, 19–20 May 2012; pp. 1936–1941.
20. Cour, T.; Benezit, F.; Shi, J. Spectral segmentation with multiscale graph decomposition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005 (CVPR 2005), San Diego, CA, USA, 20–25 June 2005; Volume 2, pp. 1124–1131.
21. Zhong, Y.; Gao, R.; Zhang, L. Multiscale and multifeature normalized cut segmentation for high spatial resolution remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6061–6075. [CrossRef]
22. Chen, J.; Li, Z.; Huang, B. Linear spectral clustering superpixel. *IEEE Trans. Image Process.* **2017**, *26*, 3317–3330. [CrossRef] [PubMed]
23. Zhu, H.; Meng, F.; Cai, J.; Lu, S. Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. *J. Vis. Commun. Image Represent.* **2016**, *34*, 12–27. [CrossRef]
24. Shen, J.; Du, Y.; Wang, W.; Li, X. Lazy random walks for superpixel segmentation. *IEEE Trans. Image Process.* **2014**, *23*, 1451–1462. [CrossRef] [PubMed]
25. 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]
26. Veksler, O.; Boykov, Y.; Mehrani, P. Superpixels and supervoxels in an energy optimization framework. In Proceedings of the European Conference on Computer Vision, Heraklion, Greece, 5–11 September 2010; pp. 211–224.
27. 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]
28. Csillik, O. Fast segmentation and classification of very high resolution remote sensing data using SLIC superpixels. *Remote Sens.* **2017**, *9*, 243. [CrossRef]
29. Moore, A.P.; Prince, S.J.; Warrell, J.; Mohammed, U.; Jones, G. Superpixel lattices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
30. 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; pp. 13–26.
31. 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.
32. Comaniciu, D.; Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [CrossRef]
33. Li, C.; Xu, C.; Gui, C.; Fox, M.D. Distance regularized level set evolution and its application to image segmentation. *IEEE Trans. Image Process.* **2010**, *19*, 3243–3254. [PubMed]
34. Zhang, X.; Cui, J.; Wang, W.; Lin, C. A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm. *Sensors* **2017**, *17*, 1474. [CrossRef] [PubMed]
35. Haralick, R.M.; Shanmugam, K. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *6*, 610–621. [CrossRef]
36. Rubner, Y.; Tomasi, C. *Perceptual Metrics for Image Database Navigation*; Springer Science & Business Media: Berlin, Germany, 2013; Volume 594.

37. Jin, X.; Davis, C.H. Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *EURASIP J. Adv. Signal Process.* **2005**, *2005*, 745309. [CrossRef]
38. Ghanea, M.; Moallem, P.; Momeni, M. Building extraction from high-resolution satellite images in urban areas: Recent methods and strategies against significant challenges. *Int. J. Remote Sens.* **2016**, *37*, 5234–5248. [CrossRef]
39. Chai, D. A Probabilistic Framework for Building Extraction From Airborne Color Image and DSM. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 948–959. [CrossRef]
40. Heuel, S.; Kolbe, T.H. Building reconstruction: The dilemma of generic versus specific models. *Künstliche Intelligenz* **2001**, *15*, 57–62.
41. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [CrossRef] [PubMed]
42. D'Angelo, P. Improving semi-global matching: Cost aggregation and confidence measure. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, B1.
43. Bafghi, Z.G.; Tian, J.; d'Angelo, P.; Reinartz, P. A New Algorithm for Void Filling in a DSM from Stereo Satellite Images in Urban Areas. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *3*, 55. [CrossRef]
44. Shahi, K.; Shafri, H.Z.; Taherzadeh, E. A novel spectral index for automatic shadow detection in urban mapping based on WorldView-2 satellite imagery. *Int. J. Comput. Electr. Autom. Control Inf. Eng.* **2014**, *8*, 1685–1688.
45. Estrada, F.J.; Jepson, A.D. Benchmarking image segmentation algorithms. *Int. J. Comput. Vis.* **2009**, *85*, 167–181. [CrossRef]
46. Liu, J.; Tang, Z.; Cui, Y.; Wu, G. Local competition-based superpixel segmentation algorithm in remote sensing. *Sensors* **2017**, *17*, 1364. [CrossRef] [PubMed]
47. Tian, J.; Krauß, T.; d'Angelo, P. Automatic Rooftop Extraction in Stereo Imagery Using Distance and Building Shape Regularized Level Set Evolution. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *42*, 393. [CrossRef]
48. Tian, J.; Cui, S.; Reinartz, P. Building change detection based on satellite stereo imagery and digital surface models. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 406–417. [CrossRef]
49. Tian, J. 3D Change Detection From High and Very High Resolution Satellite Stereo Imagery. Ph.D. Thesis, Universität Osnabrück, Osnabrück, Germany, 2013.