

Article

Classification of Ultra-High Resolution Orthophotos Combined with DSM Using a Dual Morphological Top Hat Profile

Qian Zhang ¹, Rongjun Qin ², Xin Huang ³, Yong Fang ^{1,*} and Liang Liu ¹

Received: 27 September 2015; Accepted: 1 December 2015; Published: 4 December 2015

Academic Editors: Janet Nichol, Zhong Lu and Prasad S. Thenkabail

¹ College of Electronics And Information Engineering, Sichuan University, No. 24 South Section 1, 1st Ring Road, Chengdu 610000, China; zhangqian2012@stu.scu.edu.cn (Q.Z.); liangzhai118@163.com (L.L.)

² Singapore-ETH Centre, Future Cities Laboratory, 1 CREATE Way, #06-01 CREATE Tower, Singapore 138602, Singapore; rjin@student.ethz.ch

³ School of Remote Sensing and Information Engineering, Wuhan University, No. 129 Luoyu Road, Wuhan 430079, China; huang_whu@163.com

* Correspondence: yfang@scu.edu.cn; Tel.: + 86-028-8546-3873; Fax: + 86-028-8546-3871

Abstract: New aerial sensors and platforms (e.g., unmanned aerial vehicles (UAVs)) are capable of providing ultra-high resolution remote sensing data (less than a 30-cm ground sampling distance (GSD)). This type of data is an important source for interpreting sub-building level objects; however, it has not yet been explored. The large-scale differences of urban objects, the high spectral variability and the large perspective effect bring difficulties to the design of descriptive features. Therefore, features representing the spatial information of the objects are essential for dealing with the spectral ambiguity. In this paper, we proposed a dual morphology top-hat profile (DMTHP) using both morphology reconstruction and erosion with different granularities. Due to the high dimensional feature space, we have proposed an adaptive scale selection procedure to reduce the feature dimension according to the training samples. The DMTHP is extracted from both images and Digital Surface Models (DSM) to obtain complimentary information. The random forest classifier is used to classify the features hierarchically. Quantitative experimental results on aerial images with 9-cm and UAV images with 5-cm GSD are performed. Under our experiments, improvements of 10% and 2% in overall accuracy are obtained in comparison with the well-known differential morphological profile (DMP) feature, and superior performance is observed over other tested features. Large format data with $20,000 \times 20,000$ pixels are used to perform a qualitative experiment using the proposed method, which shows its promising potential. The experiments also demonstrate that the DSM information has greatly enhanced the classification accuracy. In the best case in our experiment, it gives rise to a classification accuracy from 63.93% (spectral information only) to 94.48% (the proposed method).

Keywords: ultra-high resolution; land cover classification; digital surface models; morphology top-hat

1. Introduction

1.1. Background

Land cover classification is a well-studied, but challenging problem, which plays an important role in interpreting remote sensing data. Most of the previous investigations focus on low-to-medium resolution images for classification at the landscape level [1,2], which have been widely used for studying urban sprawl and forest mapping [3,4]. The recent development of the

very high resolution (VHR) (0.5–2 m) space-borne and aerial sensors has raised interest in the exploration of the spatial features for the land cover classification, since the spectral information alone does not constitute distinct features to separate different urban objects [5]. Some proposed methods have achieved acceptable results using VHR images for classification at the building level using 2D spatial features [6,7].

However, the new aerial sensors and platforms (e.g., unmanned aerial vehicles (UAVs)) nowadays provide remote sensing data with even higher spatial resolution (ultra-high resolution (UHR), 0.05–0.3 m). Interpreting such a kind of data is very important for urban object management and modeling [8,9], since objects that are smaller than buildings (sub-building level) become more visible and significant in UHR images, such as bus stations and cars. Due to the large perspective effects and high spectral variations, image-based spatial features may not suffice with respect to the need for the classification of UHR images. Moreover, the disparate scales of different classes are more significant.

1.2. Related Works in Classification Using Spatial Information

There are numerous works considering the improvement of classifier and feature representation for classification. Mylonas *et al.* [10] proposed to enhance the spatial representation using a fuzzy segmentation method and then adopted a voting strategy within the segments based on a pixel-based classification. Tuia *et al.* [11] proposed a multi-kernel approach that combined different kernel functions of the support vector machine (SVM) to address the different kinds of features.

Hester *et al.* [12] applied the ISODATA (iterative self-organizing data analysis technique) to the spectral bands of the VHR images and reported 89.0% overall accuracy (OA). However, in their works, ground, buildings, roofs and roads were all specified as impervious objects, which were insufficient for urban mapping applications. However, adopting the spectral information alone to distinguish these objects for finer classification will consequently result in poor results.

Researchers have demonstrated in a few studies that introducing the spatial features to describe the 2D spatial patterns of the objects brought significant improvement to the classification results [13]. Notable developments include the grey level concurrent matrix (GLCM) [14], the length-width extraction algorithm (LWEA) [15], 3D wavelet analysis [16], the differential morphological profile (DMP) [17], etc. GLCM is a set of spatial features that extract textural statistics over a given window. It was firstly proposed by Haralick *et al.* [14] for texture analysis and later recapped by Zhang [18] for building detection. Recently, it was adopted by Pacifici *et al.* [6] for classification of VHR images through a neural network, resulting in satisfactory accuracy, even with panchromatic images. LWEA was proposed by Shackelford and Davis [15], which extracted the spectrally similar area surrounding a central pixel by radiating searching lines from this centric pixel and described the spatial feature of this pixel using the longest and shortest diameter for this area. Zhang *et al.* [13] extended this representation by counting the number of pixels on the searching lines, resulting in a simplified one-dimensional spatial feature named the pixel shape index (PSI). 3D wavelet analysis was adopted as an indicator of urban complexity, as it can describe the spatial variation in the wavelet domain. A further extension from Huang and Zhang [19] considered a multi-scale approach, which varied the window size and decomposition level for more reliable spatial representation. Considering the scale problem of the VHR images and the special characteristics of the mean-shift vector [20], Qin [21] proposed a mean-shift vector-based shape feature (MSVSF), which aimed to differentiate the spatial patterns between concrete roofs and roads. Higher classification accuracy than LWEA and PSI was reported using MSVSF on the experimented dataset. Pesaresi and Benediktsson [22] computed a set of grey-scale morphological reconstruction operations on the remote sensing images by varying the size of the structural elements to build the DMP, which demonstrated effective improvements for classifying VHR remote sensing data [23].

1.3. Related Works in Classification and Data Interpretation Using Height Information

The aforementioned spatial features in Section 1.1 were proposed in a 2D context, where the 2D spatial pattern of the images is the major source of concern. A few researchers reported that integrating a third dimension (height information) could significantly improve the classification results [24,25]. This provided an interesting study path; however, the accurate and usable height data were usually expensive to obtain. The idea of using Digital Surface Model (DSM) (height information) for remote sensing interpretation has recently been popularized by the advanced development of a dense matching algorithm, which produced relatively reliable DSM from photogrammetric images. Quite a few research works were devoted to using the image-derived 3D information for change detection in 3D [26–28], which provided more reliable results. However, integrating DSM for remote sensing data classification has not been fully investigated. Only a few works have directly/indirectly reported some preliminary studies. Huang *et al.* [25] computed the GLCM feature, max-min values and variations of the DSM under a pixel-based classification framework and reported 12.3% accuracy improvement. Qin *et al.* [29] combined the orthophoto and DSM for supervised classification under a change detection framework and reported over 90% overall accuracy. Qin [30], working with others, adopted the DSM for supervised building detection, and the DSMs of a time series were used to perform the spatiotemporal inference for enhancing the building detection accuracy. A recent work from Gu *et al.* [31] proposed a multi-kernel learning model for fusing the spectral, spatial and height information, with each kernel function designed according to different features. These kernels were then linearly combined and optimized with conventional SVM for optimal performance. Most of these methods mainly considered the height information derived from LiDAR (light detection and ranging), and the performance of using image-derived DSM for UHR data was not fully evaluated.

1.4. The Proposed Spatial Feature and Object-Based Classification

In this work, we consider the problem of land cover classification for UHR remote sensing orthophotos with DSMs. As compared to the classification task in VHR data, the objects of interest in UHR data have even larger variation in scale. The increased level of detail leads to higher computational load for pixel-based classification, as well as more ambiguities in the spectral signature of the urban objects. To address these issues, we propose a dual morphological top-hat profile (DMTHP), which makes use of top-hat by reconstruction and by erosion, to extract spatial features both from the orthophoto and DSM. Considering the high dimensional feature space of the conventional morphological profiles, the sizes of the structural elements are estimated adaptively using the training data. This avoids exhaustive morphology computation of a set of sizes with regular intervals, which consequently reduces the dimensions of the feature space. A modified synergic mean-shift segmentation method is applied for object-based classification, making full use of the DSM and radiometric values of the orthophoto. In this paper, we aim to address the following issues and gaps:

- (1) There are rarely studies addressing the classification problem in ultra-high resolution detail, mainly due to the high spectral ambiguity and large perspective distortion. We incorporate 3D information to improve the traditional land cover classification problem and investigate its accuracy potential in ultra-high resolution data.
- (2) 2D spatial features are used to enhance the classification results. We aim to develop an effective and computationally-efficient spatial feature that can be applied to the 3D information, for achieving higher accuracy than traditional spatial features.
- (3) The existing research works lack quantitative evaluation on the major spatial features and their performance on the 3D information. We aim to provide such comparative studies in the course of the presentation of our novel 3D spatial feature.

The remainder of the paper is organized as follows: Section 3 introduces the proposed DMTHP features, the segmentation procedure, together with the object-based classification using the random forest (RF) classifier. In Section 3, two experiments using UHR aerial data with a 9-cm GSD (ground

sampling distance) and UAV data with a 5-cm GSD are performed and quantitative results demonstrated. A qualitative experiment is also performed over the whole Vaihingen dataset [32], which demonstrate the scalability of the proposed classification procedure. In Section 4, we validate our method by comparing it to other existing spatial features that contain height information and discuss the uncertainties, errors, accuracies and performance of our method. Section 5 concludes the paper by highlighting our contribution and the pros and cons of the proposed method.

2. Methods

2.1. Dual Morphological Top-Hat Profiles with Adaptive Scale Estimation

Mathematical morphology is regarded as a powerful tool for image processing. It was first used on binary images for shape analysis and later extended for grey-scale image analysis [22]. The opening and closing operations describe the spatial relations and provide the shape information of the image content in a local area. One of the most useful operations is morphology reconstruction, where an image J can be reconstructed as $B_{J,I}$ from a marker image I by finding the maximum of I , which is marked by $J \cdot I$ is no greater than J pixel-wise and usually derived from an erosion operation from J by a structuring element (SE) e [17]. Intuitively, the reconstruction process inherently represents the structural characteristics of the image subject to a structuring element, thus being $B_{J,\varepsilon(J,e)}$, where $\varepsilon(J,e)$ is the grey level morphology erosion:

$$\varepsilon(J,e)(i,j) = \min\{J(p-a, q-b) |, e(a,b) = 1\} \quad (1)$$

Its dual form, where J is the marker image, I is the morphological dilation using an SE e , denoted as $B_{d(J,e),J}$, where $d(J,e)$ is the grey-scale morphology dilation:

$$d(J,e)(i,j) = \max\{J(p-a, q-b) |, e(a,b) = 1\} \quad (2)$$

2.1.1. Morphological Profiles

Taking these two types of morphological reconstruction, Pesaresi and Benediktsson [17] proposed to use a set of SEs with different sizes to construct morphology profiles (MP) with a multi-scale property for remote sensing image segmentation, denoted as:

$$MP_N(J) = \{B_{J,\varepsilon(J,e_1)}, B_{J,\varepsilon(J,e_2)}, \dots, B_{J,\varepsilon(J,e_N)}, B_{d(J,e_1),J}, B_{d(J,e_2),J}, \dots, B_{d(J,e_N),J}\} \quad (3)$$

where $\{e_i\}$ is a sequence of SEs with different sizes (scales), in the form of regularly-spaced granularity. A derived form, named the differential morphological profile (DMP), is defined as the differential of the profiles:

$$\begin{aligned} \Delta_n &= B_{J,\varepsilon(J,e_n)} - B_{J,\varepsilon(J,e_{n-1})}, \overline{\Delta_n} = B_{d(J,e_n),J} - B_{d(J,e_{n-1}),J} \\ DMP_N(J) &= \{\Delta_1, \Delta_2, \dots, \Delta_N, \overline{\Delta_1}, \overline{\Delta_2}, \dots, \overline{\Delta_N}\} \end{aligned} \quad (4)$$

This is effective to sense the structural differences of the image content at different scales. MPs and DMPs were experimented with by Benediktsson *et al.* [23] in the classification of high resolution remote sensing images of the urban areas. Tuia *et al.* [33] examined a set of morphological operators on classifying panchromatic images and demonstrated the effectiveness of the MPs. Each set of operators may delineate particular classes, but they could not fully take advantage of the classes and compensate for each other. The morphological reconstruction avoids discontinuities, and this retains redundancies across different scales in the MPs.

2.1.2. Morphological Top-Hat Profiles

Morphological top-hat (MTH) is defined as the peaks of an image grid, computed by morphological operations. Intuitively, the top-hats can detect the spatial blobs of the image grid. Such spatial blobs can be used effectively to represent urban structures, such as buildings, cars and shadows cast by buildings. Huang and Zhang [34,35] applied the multi-directional morphological

top-hat transform to detect buildings and shadows in panchromatic images. They assumed that buildings dominate the bright region and that the shadows cast by the buildings are revealed as dark blobs. Qin and Fang [36] indicated that the height of a blob is a better index for representing buildings. They adopted the top-hats on the DSMs and combined with NDVI (normalized difference vegetation index) filtering to perform accurate building detection.

The success of top-hats in detecting urban objects inspired the idea to build morphological top-hat profiles to deal with multi-scale urban objects in UHR remote sensing data. We consider two types of morphological top-hats, (1) top-hat by reconstruction (THR) and (2) top-hat by erosion (THE), which can be simply defined as follows:

A. Top-hat by reconstruction:

$$THR(J, e) = J - B_{J, \varepsilon(J, e)} \quad (5)$$

B. Top-hat by erosion:

$$THE(J, e) = J - \varepsilon(J, e) \quad (6)$$

THR is effective at detecting the peaks of an image grid. However, one drawback of the THR is that it cannot highlight off-terrain objects on a slanted surface connecting the top of the blobs, since the morphological reconstruction process finds the peaks globally over the image grids. Qin and Fang [36] partly addressed this problem by blocking the connections made by vegetation using NDVI before THR. On the contrary, THE simply detects the local height extreme subject to the given SE. It can detect the off-terrain objects on a slanted surface, whereas it produces errors for terrain objects. Figure 1 shows an example, where THR and THE are computed using a disk-shaped SE. The results of THR (Figure 1c) can effectively detect the off-terrain objects and, at the same time, maintain good separability of the terrain classes (road, ground). THE is able to highlight the local maximum under the area marked by the SE, but results in errors in the terrain objects (e.g., roads). The area in Figure 1 marked with a red circle shows a building that connects to the adjacent road, which has a similar height. The THR (shown in Figure 1c) misses a part of the buildings, while THE has detected this part.

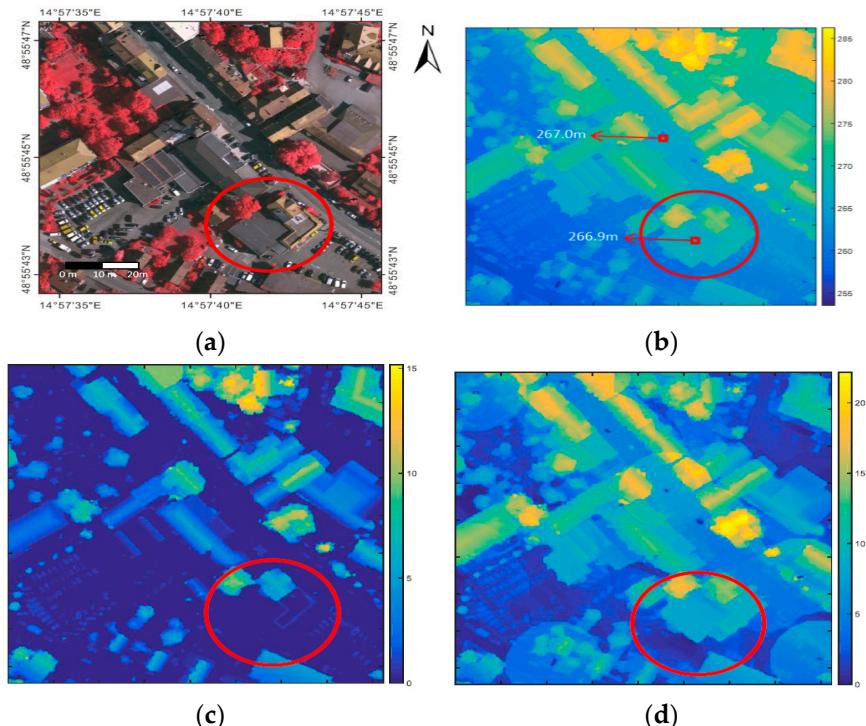


Figure 1. Results of top-hat by reconstruction (THR) and top-hat by erosion (THE) on a DSM of a test area (meters). (a) The orthophoto; (b) DSM; (c) THR; (d) THE. (The marked area is explained in the text).

These two top-hats can be used to compensate for each other. As the sizes of the urban objects vary a lot, we consider a multi-scale approach. A series of SEs $\{e_i\}$ with different sizes are used to construct the DMTHP:

$$DMTHP(J)_N = \{THR(J, e_1), THR(J, e_2), \dots, THR(J, e_N), THE(J, e_1), THE(J, e_2), \dots, THE(J, e_N)\} \quad (7)$$

2.1.3. Adaptive Scale Estimation

The multi-scale SEs are effective at describing the spatial differences of objects with different sizes. However, as indicated by Benediktsson *et al.* [23], a major drawback of such a strategy is the high computational cost for classification with high dimensional features. To reduce the high dimension, they only use two elements with the maximal responses for training and classification, but the computation of the morphological reconstruction still needs to be done at full scale, which is considerably time consuming. Moreover, the morphological top-hats may contain redundancies, since the results are closely related to the scale of the objects.

Figure 2 shows an example of the THR and THE profiles of different classes. It can be seen that the profiles in each sample of the urban object show different patterns. However, large redundancies can be observed, both in the profile bar and the THR and THE maps in the last two rows. The THR and THE maps in the last three columns remain similar. In the bar figure of the car, some values at different scales stay the same. This is due to the fact that the scales of different urban classes are within a certain range, and a fixed interval of scale may not contribute effectively to the distinct features for classification. Only the values near the discontinuities are useful to construct distinct features, corresponding to the scale of different urban objects. Since for most of the classification tasks, the training samples are selected as representatives of the urban object, it is feasible to estimate the scale bounds of different urban classes based on the training samples.

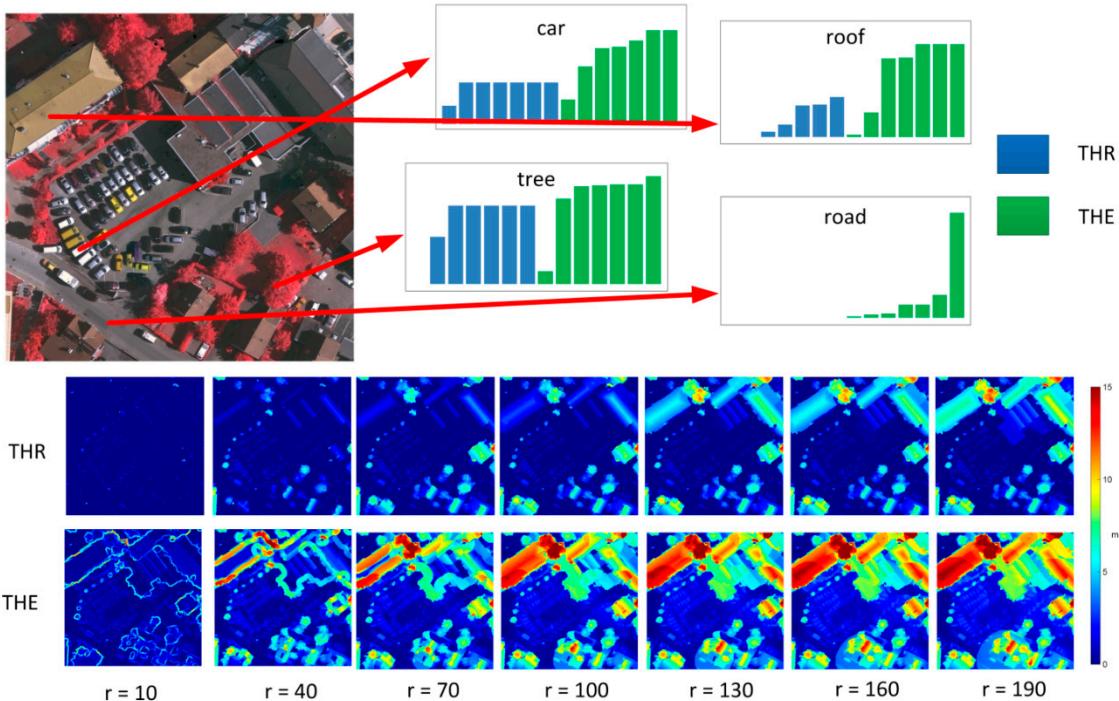


Figure 2. The THR and THE values at different scales for difference urban classes in the ultra-high resolution (UHR) image. “r” is the size of the structuring element (SE) in pixels.

Given a set of training samples for N classes $\{\mathcal{C}_1\}_{n_1}, \{\mathcal{C}_2\}_{n_2}, \dots, \{\mathcal{C}_N\}_{n_N}$, where $c_i^m \in \{\mathcal{C}_m\}$ is a segment in the image space specifying a sample for class \mathcal{C}_m , we compute its scale $S(c_i^m)$ as:

$$S(c_i^m) = \sqrt{R_x(c_i^m)^2 + R_y(c_i^m)^2} \quad (9)$$

where $R_x(\cdot)$ and $R_y(\cdot)$ are the ranges of the segment in the x and y directions, respectively. We compute the upper bounds of the scales in each class as:

$$\text{UB}(\mathcal{C}_m) = \max\{\mathcal{S}(c_i^m), c_i^m \in \{\mathcal{C}_m\}\} \quad (9)$$

The $\text{UB}(\mathcal{C}_m)$ are sorted, and adjacent scales whose distance to the others is smaller than a threshold $\tau = 80$ pixels are clustered. The final estimated scales are denoted as $\{r\}_k$. Thereby, the DMTHP can be reformulated as:

$$\begin{aligned} \text{DMTHP}(J)_k \\ = \{ & \text{THR}(J, e_{r_1}), \text{THR}(J, e_{r_2}), \dots, \text{THR}(J, e_{r_k}), \text{THE}(J, e_{r_1}), \text{THE}(J, e_{r_2}), \dots, \text{THE}(J, e_{r_k}) \} \end{aligned} \quad (10)$$

In our experiment, to ensure a numerically-equivalent contribution for each element of the feature, we normalize each dimension of the feature to [0,1] across the whole dataset.

2.2. Spectral- and Height-Assisted Segmentation for Object-Based Classification

2.2.1. Height-Assisted Synergic Mean-Shift Segmentation

The UHR data reveal a high level of detail of the ground objects. Therefore, it is necessary to adopt object-based analysis to reduce the computational complexity. To make full use of the spectral and height information, we adopt a height-assisted segmentation that employs both the DSM and color images, as proposed in Qin *et al.* [29]. The segmentation method is essentially a synergic mean-shift (MS) segmentation [20,37], which applies MS segmentation and, at the same time, constrains the segment boundary using a weight map, implying the probability of each pixel being an object boundary. In this method, the weight map is defined as the Canny magnitude [38] of the DSM. In each iteration of the segmentation, the synergic segmentation prevents the MS procedure from going beyond pixels with a high boundary probability. In the classic MS segmentation, there are two major parameters, the spatial bandwidth H_s and the spectral bandwidth H_r , which have spatial proximity and spectral similarity for the segmentation procedure [20]. In addition to the classic MS segmentation, the synergic MS segmentation has another parameter β that controls the weight of the additional edge constraint, which in our context is derived from the DSM.

Figure 3 shows an example of the synergic MS segmentation. It can be seen in the area outlined by the red circle that the height-assisted synergic MS segmentation is able to break the segments with height jumps. This provides more accurate segments for further training and classification. In our experiments, $H_s = 7$, $H_r = 4$ and $\beta = 0.1$ are set as constants, as suggested in [29]. Due to the DSM constraint, $H_r = 4$ is set as a relatively large value to reduce the effect of over-segmentation.



Figure 3. An example of synergic mean-shift (MS) segmentation. (**Left**) Classic mean-shift segmentation; (**middle**) height-assisted synergic MS segmentation; (**right**) boundary probability map derived from the DSM. The interpretation of the red circle is in the text.

2.2.2. Classification Combining the Spectral and DMTHP Features

The random forest (RF) classifier [39] is widely used for classifying features with hierarchical characteristics. Qin *et al.* [29] have demonstrated that RF performs better than support vector

machine (SVM) when combining the height and spectral information for classification, since they do not linearly contribute to the final feature vector. In our experiment, RF is applied to classify the feature vectors constituted by the spectral and DMTHP features. RF is essentially an ensemble learning method using a decision tree classifier. The advantages of this method are the improved accuracy due to the voting strategy of multiple decision trees and the hierarchical examination of the feature elements, which are particularly useful for features constructed from different sources.

Since the spectral feature is the major driving force for the classification, we adopt the principal component analysis (PCA) transformation of the image color bands as the spectral information, since it has been proven to give better classification accuracy [13]. It maximizes the variance of the spectral direction to increase the independence of each band. The DMTHP feature is applied to both the orthophoto and DSM, as the morphological top-hats on both data sources return useful information, as reported by Huang and Zhang [35] and Qin and Fang [36]. For the orthophoto, we apply the DMTHP to the brightness and darkness images, which are defined as the first component of the PCA transform and its inverse image, being effective for describing bright blobs and dark blobs at different scales. For the height information, the DMTHP is directly applied to the DSM. We finally concatenate these features into a vector-stack fashion to perform the random forest classification. This feature extraction procedure is described in Figure 4.

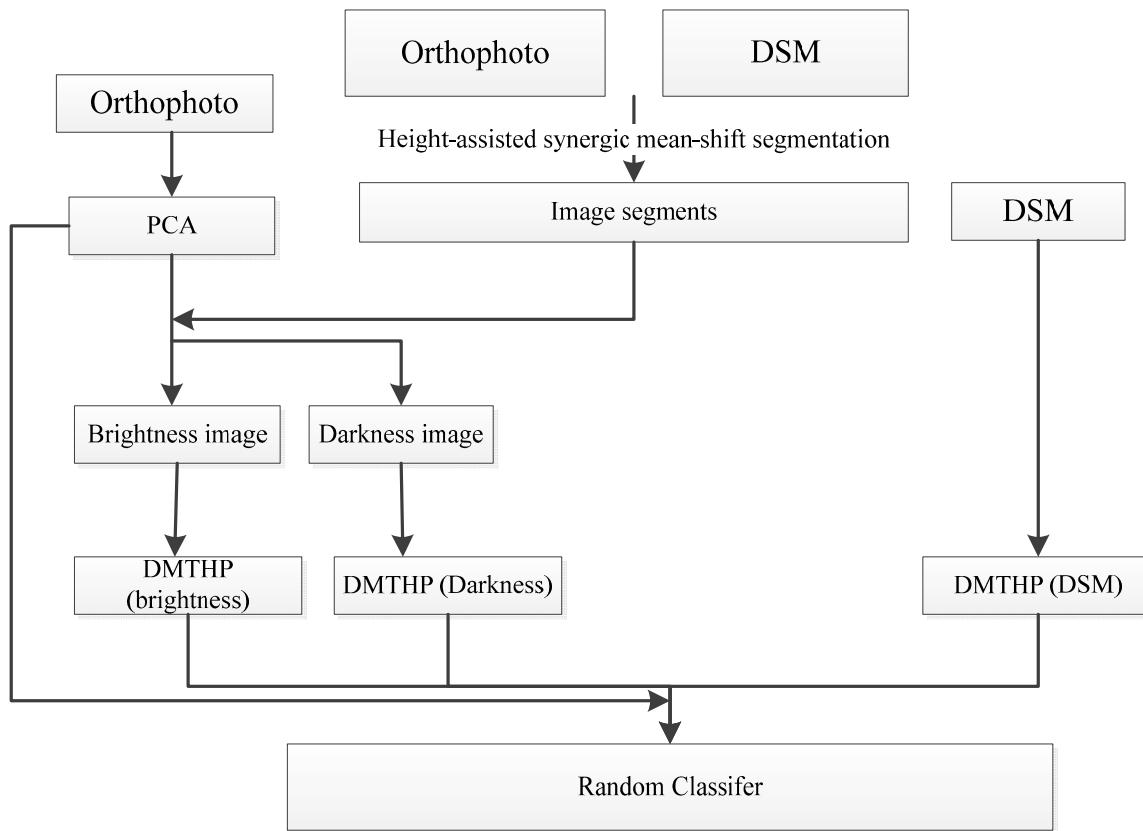


Figure 4. The proposed workflow for UHR image land cover classification. DMTHP, dual morphology top-hat profile.

3. Experimental Results

3.1. Experimental Setup

Our experiments mainly target the UHR remote sensing data, and the purpose of the experiments is to (1) test our proposed feature on spatially-varying urban objects and (2) evaluate

the accuracy potential of UHR data using the classic land cover classification paradigm. This section contains the three experiments on aerial and UAV images, with associated DSM derived from image dense matching techniques. The first experiment is performed on a small test area from the Vaihingen dataset [32], with a 9-cm GSD. The second experiment is performed on a dataset generated using UAV images, with a GSD of 5 cm. Both Experiments 1 and 2 are quantitatively evaluated against the ground truth. The third experiment applies the proposed classification flow to the whole Vaihingen dataset to qualitatively demonstrate the scalability of the classification procedure.

The RF classifier is adopted for the classification, and 500 decision trees are used for the training procedure; the number of variables for classification is computed as the square root of the feature dimension. In our experiment, the feature dimension is 12 and 15 for Experiments 1 and 2. A large part of the test datasets is manually labeled as reference data. Around 1 percent of the marked labels are randomly selected (as shown in Table 1), and a 5-fold cross-validation (CV) process is applied to ensure the statistical robustness of the proposed features, by eliminating the possibility of sample-independent results. The number of training samples and test samples is listed in Table 1.

Table 1. Statistics of the training and test samples.

		Experiment 1	Experiment 2
Training sample for each class	Building	51	101
	Road	72	103
	Tree	51	101
	Car	53	110
	Grass	52	103
	Ground	53	/
	Shadow	53	/
	Water	/	118
Total training samples		385	636
Total test samples		10,312	32,947
Total segments		41,465	46,867
Percentage (%)		0.93	1.35

3.2. Experiment with Test Dataset 1

The first test dataset and the resulting classification map are shown in Figure 5. It can be seen that the scene is composed of rather complex objects, with buildings varying in shapes and sizes and cars distributed in different areas with different densities. The orthophoto and DSM are derived from the aerial images by the INPHO 5.3 software, using multiple feature matching. The zoomed-in image (Figure 5c) shows the ground objects in a very high level of detail, and the color-coded DSM shows that the relief differences of the small parked cars are well captured. The false-color orthophoto (shown in Figure 5a) contains a near-infrared band, which is effective for distinguishing vegetation and concrete. Therefore, both the spectrum and DSM constitute rather informative sources to classify the complicated and detail scene. Table 2 shows the statistics of the classification against the ground truth. The proposed method has achieved 93% overall accuracy. The buildings and roads are well distinguished from each other and have over 90% classification accuracy. More than 80% of cars are correctly classified in the classification map.

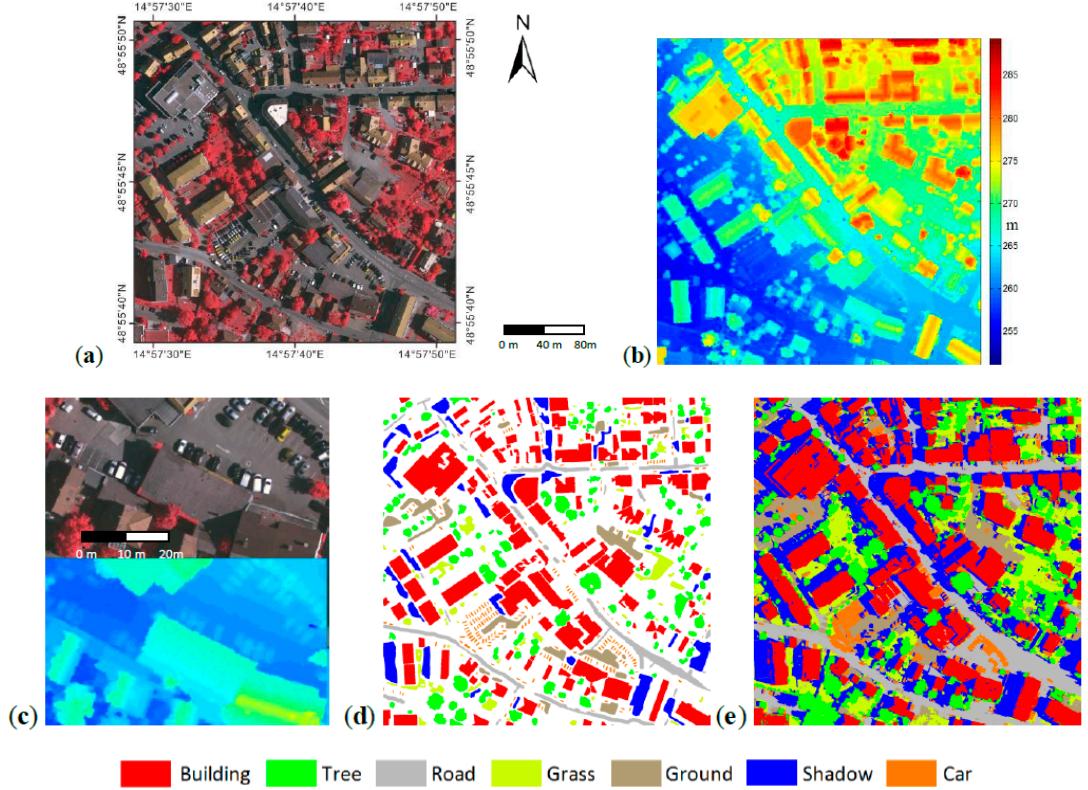


Figure 5. (a) The orthophoto; (b) DSM; (c) zoom-images of a parking area; (d) reference data; (e) classification map of the proposed method.

Table 2. Classification results of Experiment 1 (percentage (%)) (per-class producer's (P) and user's (U) accuracy). CV, cross-validation.

	Building	Road	Tree	Car	Grass	Ground	Shadow	CV	OA
(P)	94.77	95.97	96.22	83.04	99.33	72.20	99.23		
(U)	98.66	75.87	98.37	98.88	96.90	93.15	85.81	88.49	93.98

3.3. Experiment with Test Dataset 2

We include a second dataset from a different data source, as shown in Figure 6. The orthophoto and DSM are generated from a UAV mission, as described in [9,40], with a 5-cm GSD. The DSM is generated using a hierarchical semi-global matching algorithm [41,42]. The slightly increased resolution of the dataset brings some challenges. A major challenge of this dataset is that the colors of the roof and the ground are very similar to each other, with the roads directly connected with the parking lots. Therefore, in this experiment, we consider such ground as roads. Moreover, the roof tops take over a large part of the image content, which have a similar or even larger area than the ground, and this might create potential problems for the sequence of the MTH to identify the correct object class.

Figure 6d and Table 3 show the resulting classification map and the associated statistics. We have obtained more than 94% overall accuracy. Over 96% of buildings and roads are identified well, and 83% of the cars in the parking lots are correctly detected, which could facilitate the application of urban infrastructure management. The classification accuracy of the vegetation is lower than Experiment 1, and this is expected, since we do not have the near-infrared information in this experiment.

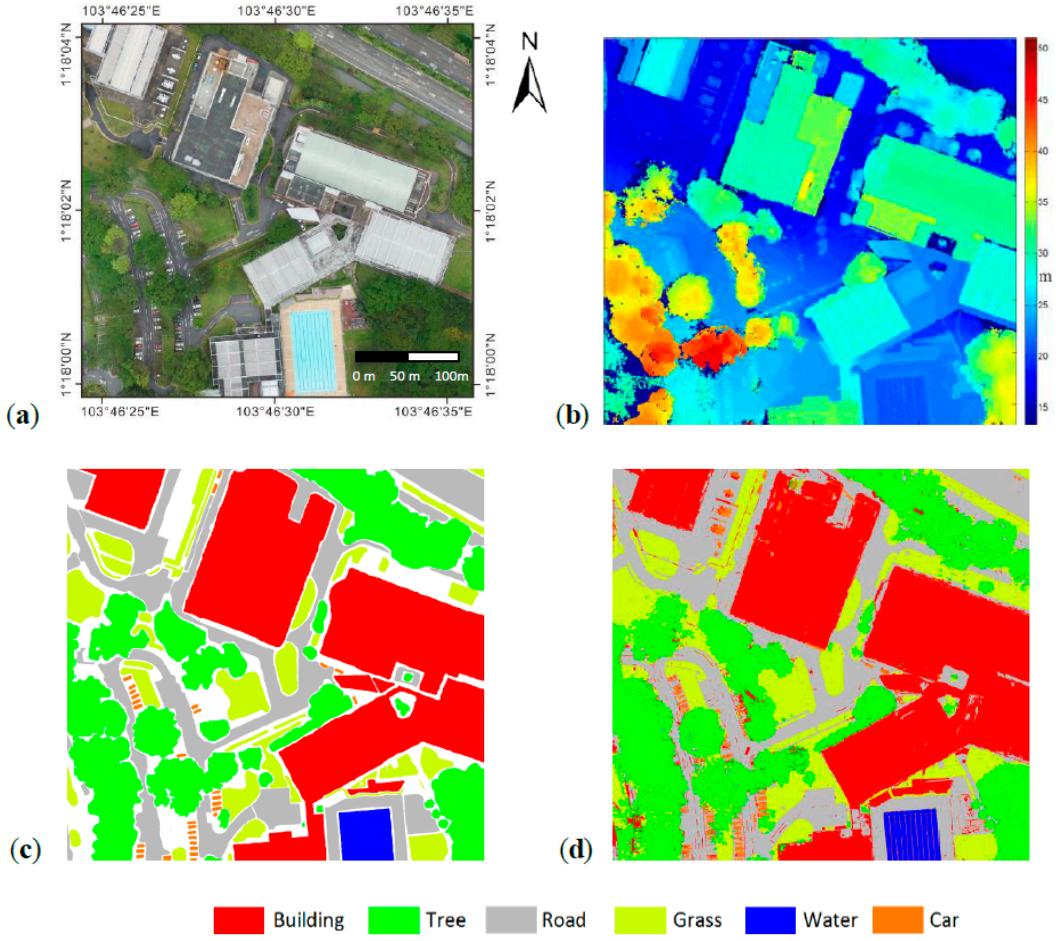


Figure 6. (a) The orthophoto; (b) DSM generated from the UAV images; (c) reference data; (d) classification with the proposed method.

Table 3. Classification results of Experiment 2 (percentage (%)) (per-class producer's (P) and user's (U) accuracy).

	Building	Road	Tree	Car	Grass	Water	CV	OA
(P)	96.03	96.30	90.33	83.04	79.63	91.95		
(U)	89.26	85.71	97.59	89.86	90.91	99.99	89.15	94.48

3.4. Test Dataset 3

The quantitative experiments have shown the advantages of our proposed features against the tested methods. However, it is crucial to test the proposed method over a large dataset. In this experiment, we applied our method to the whole Vaihingen dataset, which is $20,000 \times 20,000$ in its dimensions. However, due to the lack of reference data for such a highly-detailed classification task, we only evaluate the whole dataset qualitatively by visual comparison, and the quantitative study of this whole dataset is still in progress. The training samples are around 0.1 percent of the total number of segments. We have selected two representative areas, as shown in Figure 7. The visual comparison of these two areas to the orthophoto has shown that the classification maps have described the scene well, including small plants and vehicles. There is a small part of the misclassifications occurring on the road in the upper left part of the scene, classifying the road as cars. This might be caused by sampling selection and the image matching errors.

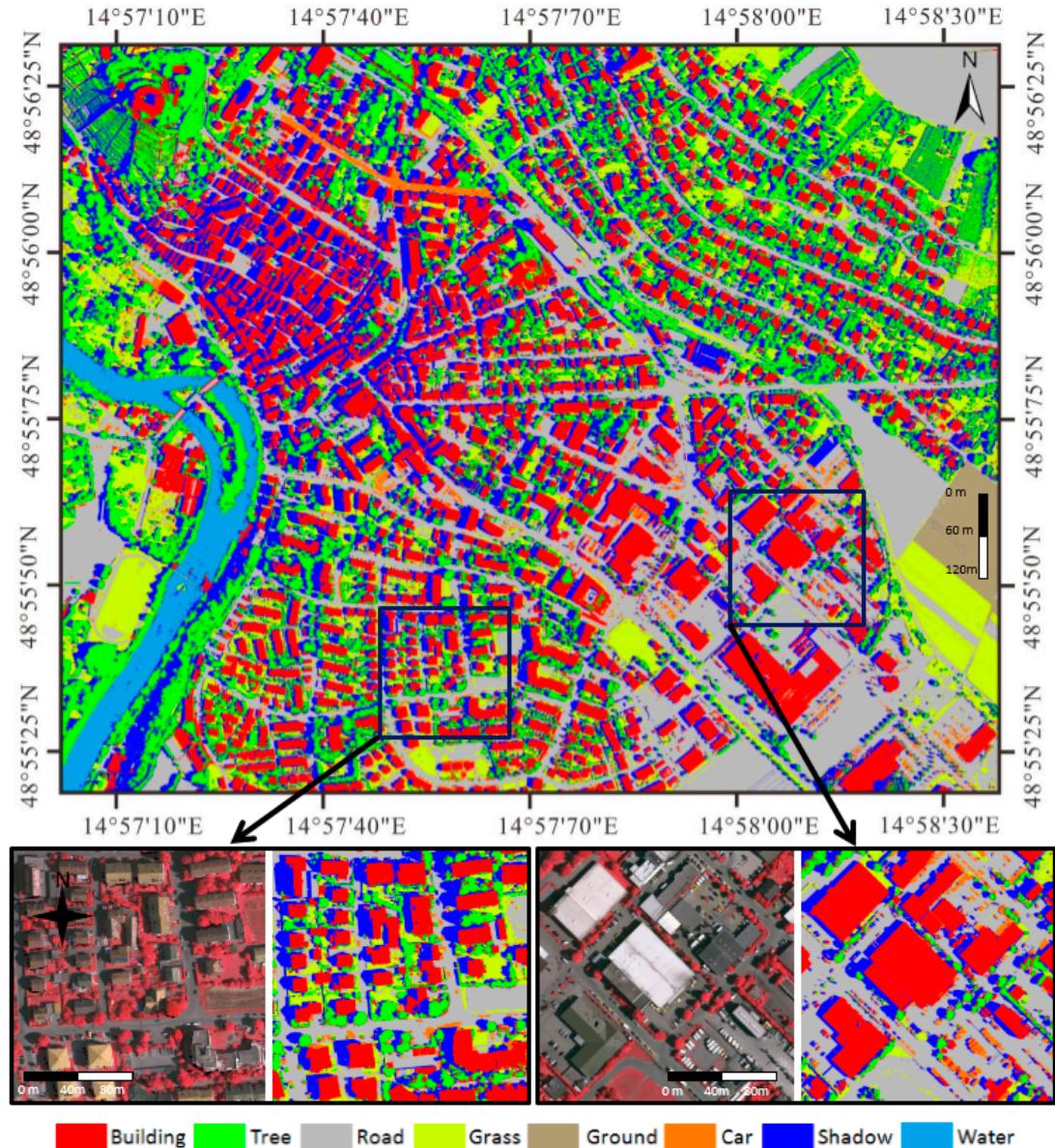


Figure 7. Qualitative experimental results the whole Vaihingen dataset. Two zoom-in areas are shown for a visual comparison.

4. Validations and Discussions

4.1. Comparative Studies and Validations

To illustrate the effectiveness of the proposed DMTHP feature, we compare it to other spatial features with height information, and these include DMP, mean height information and nDSM (normalized DSM). The DMP is applied in the same manner as the DMTHP feature described in Figure 4, which computes the differential morphology profiles of the brightness image, the darkness image and the DSM. The size of the structuring elements ranges from 10 pixels–300 pixels, with a regular interval (this interval is fixed as 30 pixels in our experiment), which results in a 30-element spatial feature. The nDSM is very effective at representing the off-terrain objects, usually computed by subtracting the DTM (digital terrain model) from the DSM. Since we do not have a separate DTM for this area, we adopt the morphological top-hat by reconstruction as an estimation, which has been widely used as an approximation of the nDSM [36]. The comparative studies of Experiments 1 and 2 are quantitatively performed, with the resulting classification maps and associated accuracies presented in Figures 8 and 9 and Tables 4 and 5.

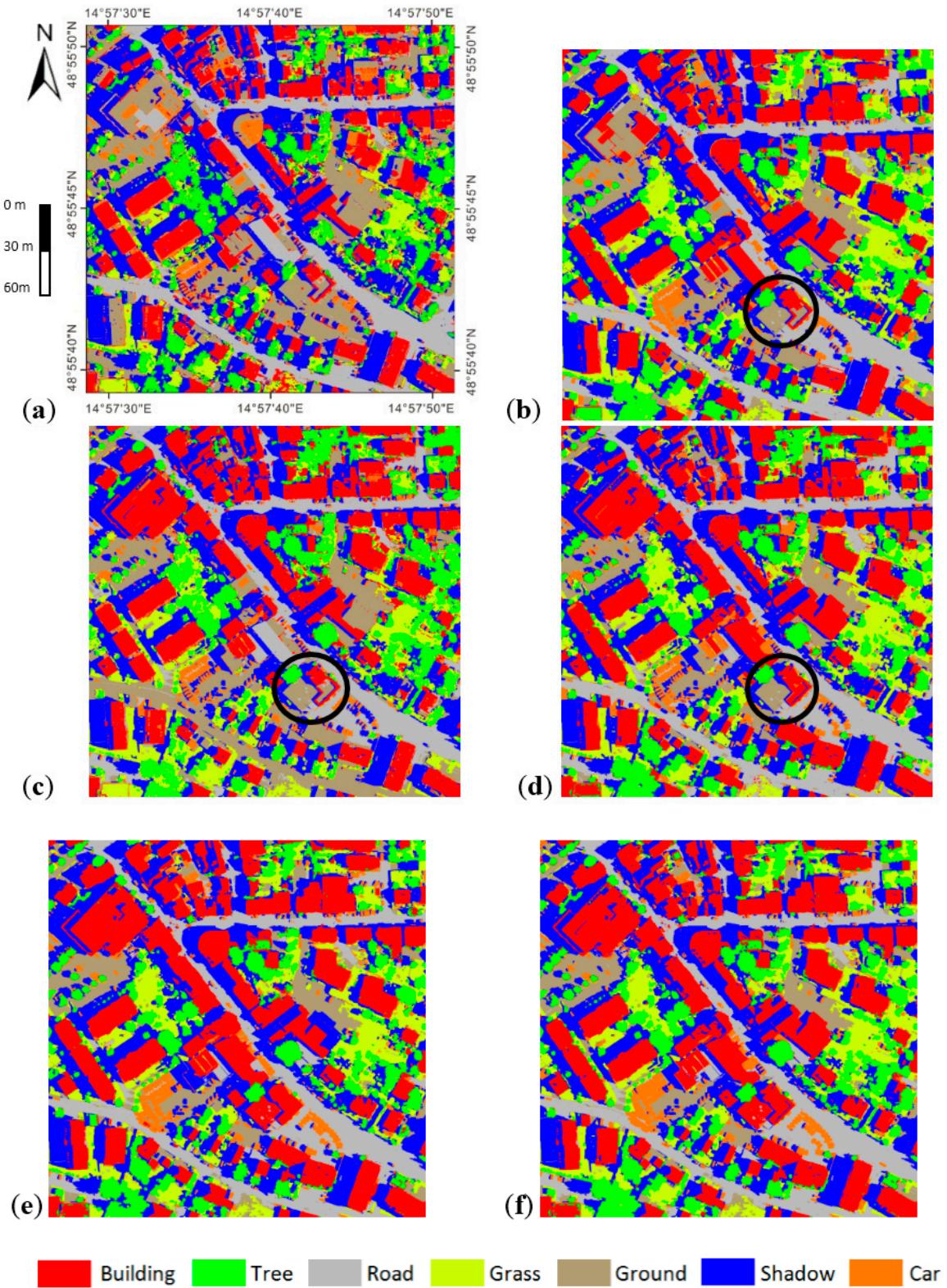


Figure 8. The classification maps of Test Dataset 1 using PCA combined with different spatial features. **(a)** Using spectral features only (PCA); **(b)** PCA + DMP; **(c)** PCA + height; **(d)** PCA + normalized DSM (nDSM); **(e)** PCA + DMTHP with a regular interval radius sequence; **(f)** PCA + DMTHP with adaptive radius selection (the black circles are explained in the text).

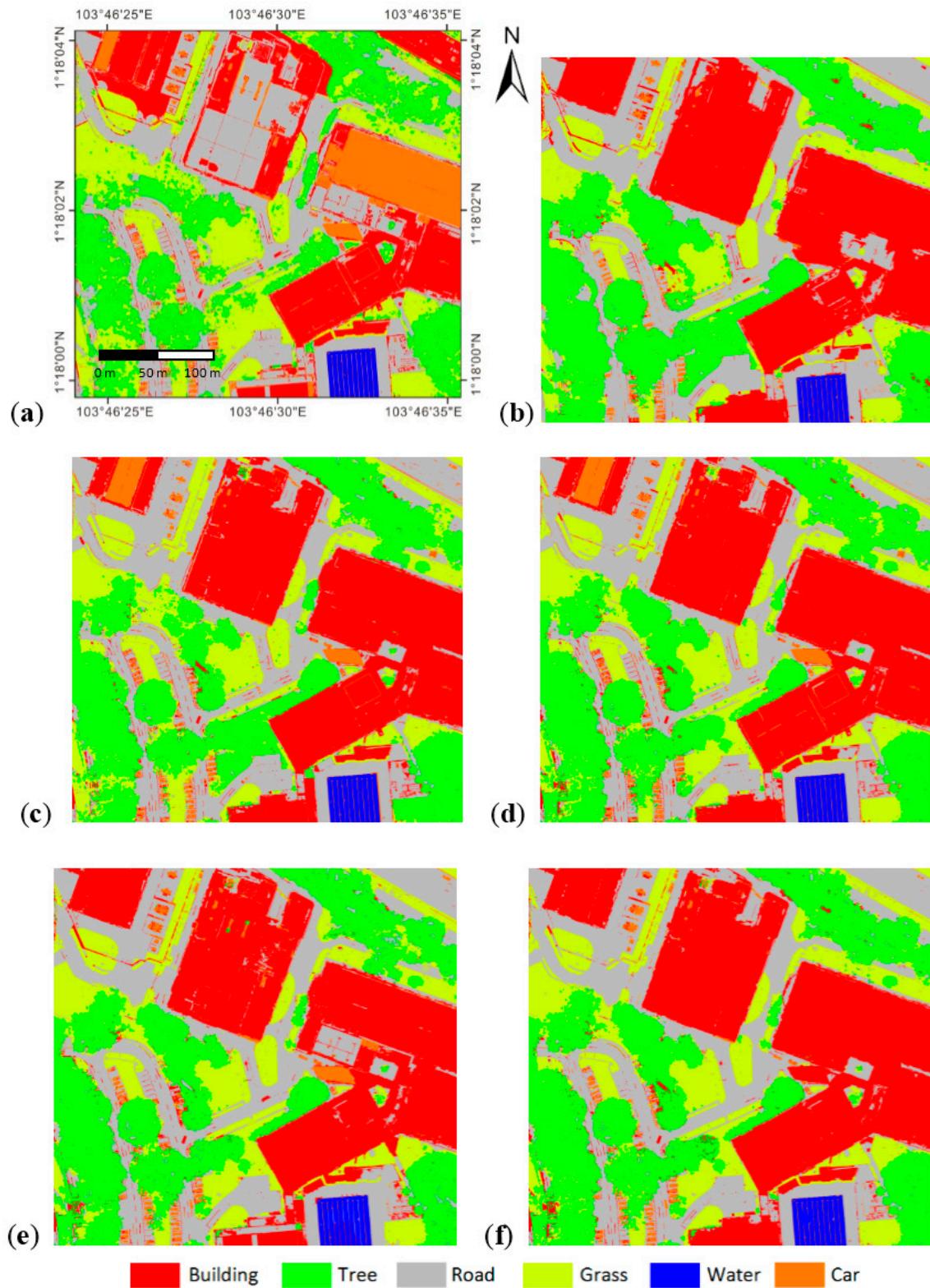


Figure 9. The classification maps of Test Dataset 2 using PCA combined with different spatial features. (a) Using spectral features only (PCA); (b) PCA + DMP; (c) PCA + height; (d) PCA + nDSM; (e) PCA + DMTHP with a regular interval radius sequence; (f) PCA + DMTHP with adaptive radius selection.

Table 4. Comparative Results of Experiment 1 (per-class producer's (P) and user's (U) accuracy).

%		PCA	PCA + DMP	PCA + Height	PCA + nDSM	PCA + DMTHP (Regular)	PCA + DMTHP (Adaptive)
Building	(P)	60.74	76.98	77.36	89.53	92.19	94.77
	(U)	88.18	95.48	95.21	95.15	96.17	98.66
Road	(P)	89.28	96.29	67.25	96.62	96.29	95.97
	(U)	78.99	76.24	72.21	75.56	76.06	75.87
Tree	(P)	70.04	88.36	84.67	90.41	94.00	96.22
	(U)	81.38	96.90	84.68	96.25	97.99	98.37
Car	(P)	64.87	73.41	69.08	70.03	78.57	83.04
	(U)	93.19	94.73	96.71	95.77	98.85	98.88
Grass	(P)	83.43	99.03	84.22	97.34	99.16	99.33
	(U)	75.92	91.42	84.00	91.55	96.35	96.90
Ground	(P)	90.19	72.22	80.35	72.18	72.22	72.20
	(U)	70.37	76.80	61.66	87.76	93.26	93.15
Shadow	(P)	92.35	99.37	98.09	98.22	99.40	99.23
	(U)	74.27	79.83	79.35	80.23	81.61	85.81
CV		71.58	85.08	80.49	81.94	87.62	88.49
OA		71.50	83.61	80.29	89.98	92.27	93.98

Table 5. Comparative Results of Experiment 2 (per-class producer's (P) and user's (U) accuracy).

%		PCA	DMP	PCA + Height	PCA + nDSM	PCA + DMTHP (regular)	PCA + DMTHP (adaptive)
Building	(P)	47.96	93.60	91.32	90.29	89.68	96.03
	(U)	58.90	87.23	83.46	81.15	85.42	89.26
Road	(P)	77.65	95.82	94.19	94.43	95.60	96.30
	(U)	62.44	79.17	83.22	85.55	79.25	85.71
Tree	(P)	61.37	91.06	84.78	89.10	88.15	90.33
	(U)	89.25	87.27	82.71	96.68	97.86	97.59
Car	(P)	73.40	76.20	70.44	71.22	79.32	79.63
	(U)	68.58	88.07	82.41	82.93	92.31	89.86
Grass	(P)	93.43	86.30	81.25	96.95	96.96	97.21
	(U)	71.73	90.29	82.51	89.22	90.19	90.91
Water	(P)	90.96	90.88	90.97	91.66	90.88	91.95
	(U)	99.99	99.99	99.99	99.99	99.99	99.99
CV		63.85	86.60	83.87	85.51	86.41	89.15
OA		63.93	92.40	88.87	91.53	91.70	94.48

Figure 8 shows the classification results using four different features associated with DSM, and Table 4 lists their classification accuracies, where CV is positively correlated with the OA. It can be seen that the classification result using spectral information alone (Figure 6a) produces many classification errors, which mainly occurs in the building and road/ground classes, as these classes have very similar spectral information. The incorporation of DSM renders better results. However, there are still misclassifications between the roof tops and the ground (e.g., shown in the black circle in Figure 8b–d). nDSM and DMP have obtained better accuracy, as they reveal the spatial structure of the image. Among the tested features, the proposed DMTHP with adaptive radius selection achieved the best OA, while the OA of DMTHP with a regular interval radius is slightly lower, as the adaptive radius selection procedure can avoid redundant information when constructing the feature vector.

Figure 9 shows the resulting classification maps, and Table 5 lists the associated classification accuracies. It is expected that we observe low classification when using the spectral information

alone (Figure 8a), as a large roof top is classified as the concrete ground. The DMP in this experiment produces better OA than the nDSM. The DMTHP with adaptive radius selection has obtained the best OA, 2.78% higher than the regular interval DMTHP. The large improvement is due to the fact that our sequence radius is fixed between 10 and 300, while the adaptive radius selection procedure correctly estimated the granularity of the building roofs and the other objects.

4.2. Uncertainties, Errors, Accuracies and Performance

Both of the comparative studies have demonstrated that the consideration of DSM information generally has significantly increased the classification accuracy, and our proposed DMTHP feature outperforms the tested methods. In the best case, as shown in Table 5, our proposed method has obtained an overall accuracy of 94%, in comparison with the traditional method that only adopts spectral information (64%). In this case, the resulting accuracy can suffice with respect to the general need for accurate land cover mapping and object identification.

The incorporation of DSM in the classification procedure using our proposed DMTHP feature has obtained satisfactory results. However, being that there is a problem with the classification tasks, the major uncertainties that affect the resulting accuracy are the selection of samples, as well as the data quality. In our case, the accuracy of the DSM is a critical factor. For small objects, such as cars and public facilities (benches on the street), the matching algorithm may fail to capture their elevation, which consequently causes a wrongly-described DMTHP feature on these objects. Therefore, the selection of the samples should be random, but also considerate of such a matching failure.

The proposed DMTHP feature is based on morphology top-hat reconstruction, which may require high computational load in terms of feature extraction. The adaptive scale selection strategy proposed in Section 2.1.3 can effectively reduce the computational time. In particular, we have recorded the running time of the DMTHP with and without adaptive scale selection, which is shown in Table 6.

Table 6. Running time of the different steps of the classification. DMP, differential morphological profile.

Time (s)	Experiment 1			Experiment 2		
	DMP	DMTHP (Regular)	DMTHP (Adaptive)	DMP	DMTHP (Regular)	DMTHP (Adaptive)
Segmentation	471.20	468.54	468.54	821.51	821.25	821.25
Feature Extraction	396.08	260.57	71.87	551.48	360.83	457.37
Training	1.31	7.44	0.68	2.38	1.18	0.76
Classification	1.28	2.03	0.81	1.44	1.15	1.12
Total	869.87	738.59	541.96	1376.81	1184.41	1280.50

The implementation is done with a normal PC, with the mixed use of MATLAB and the C++ programming language. The running time comparison in Table 6 is performed under the same conditions. In Experiment 1, it shows that the adaptive DMTHP has a significantly shorter running time in the feature extraction, training and classification stages, which mainly is attributed to the reduced feature dimensions. The overall running time of the adaptive DMTHP is shorter than the others, while the feature extraction time is longer, which may be caused by a larger radius estimated by our adaptive radius selection procedure.

5. Conclusions

In this work, we have proposed a dual morphological top-hat profile (DMTHP), which extracts spatial features from the orthophoto and DSM. A simple and adaptive scale selection strategy that determines the granularity sequences of the profiles is suggested to effectively reduce the computational cost, as well as the dimensionality of the feature. We have further applied the proposed feature to address the problem of the land cover classification on UHR remote sensing

images combined with the associated DSM, aiming to interpreting urban objects at a sub-building level (such as cars). The random forest classifier was adopted under an object-based scenario, in which the segmentation was performed using the advanced synergic mean-shift algorithm, combining both images and DSM.

The proposed method has been compared to several spatial features, and this demonstrates that the proposed DMTHP together with the adaptive scale strategy obtained the best OA. The DMTHP with adaptive radius selection has obtained 10% and 2% improvement in the OA compared to the well-known DMP feature in our two quantitative experiments. As for the comparison to the classification using merely the orthophoto, the incorporation of the DSM has demonstrated a notable improvement. In the best cases in our experiment, it increases the OA from 63.93% (spectral information only) to 94.48% (adaptive DMTHP).

A qualitative experiment using the whole Vaihingen dataset ($20,000 \times 20,000$) has shown the possibility of the proposed method applied to large-scale datasets, and the visual comparison shows the promising potential of the proposed method.

In general, our contribution in this paper lies in the following aspects:

- (1) We have presented a novel feature DMTHP with adaptive scale selection to address large-scale variation of urban objects in the UHR data, as well as reduced the computational load and feature dimensionality, which have obtained the optimal classification accuracy in comparison with existing features (2%–10% enhancement to the well-known DMP feature and other height features).
- (2) We have demonstrated that in the best case, the proposed method has improved the classification accuracy to 94%, as compared to 64% using only spectral information. This is important to draw the attention of the land cover mappers to consider the use of the height information for land cover classification tasks.
- (3) A complete quantitative analysis of different UHR data with a 9-cm and a 5-cm GSD has been performed, with comparative studies on some of the existing height features. This provides valid insights for researchers working on 3D spatial features.
- (4) We have performed a qualitative experiment with $20,000 \times 20,000$ pixels, which has shown that the proposed method can be used in a large-scale dataset to obtain very detailed land cover information.

Our experiments show that it is feasible to combine the orthophoto and DSM for urban object classification to obtain satisfactory accuracy. However, there are still some misclassifications in our experiments. These misclassifications mainly occur on large roofs with a similar color to the ground, as well as the underrepresented sample selection (as for the classification of a large scene, such as our qualitative experiment). Some errors are due to the image matching errors, which fail to capture the elevation differences of the cars to the ground. Therefore, our future work will focus on a better fusion of the spectral and height information. Since our current version of morphological top-hat is grey level based, morphological top-hat profiles based on color blobs will be considered to address these problems.

Acknowledgements: This work was jointly established at the College of Electronics And Information Engineering, Sichuan University and the Singapore-ETH Centre for Global Environmental Sustainability (SEC), co-funded by the Singapore National Research Foundation (NRF) and ETH Zurich. The Vaihingen dataset was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) (Cramer, 2010), URL: <http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html>.

Author Contributions: All authors contributed to this paper. Rongjun Qin and Qian Zhang proposed the idea. Qian Zhang processed the data, performed the experiment and prepared the figures and tables. Rongjun Qin and Qian Zhang wrote the paper. Xin Huang provided valuable comments and ideas and proofread the paper. Yong Fang and Liang Liu provided suggestions that improved the quality of the paper.

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

References

- Tucker, C.J.; Townshend, J.R.; Goff, T.E. African land-cover classification using satellite data. *Science* **1985**, *227*, 369–375.
- De Fries, R.; Hansen, M.; Townshend, J.; Sohlberg, R. Global land cover classifications at 8 km spatial resolution: The use of training data derived from Landsat imagery in decision tree classifiers. *Int. J. Remote Sens.* **1998**, *19*, 3141–3168.
- Collins, J.B.; Woodcock, C.E. An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. *Remote Sens. Environ.* **1996**, *56*, 66–77.
- Yuan, F.; Sawaya, K.E.; Loeffelholz, B.C.; Bauer, M.E. Land cover classification and change analysis of the twin cities (Minnesota) metropolitan area by multitemporal Landsat remote sensing. *Remote Sens. Environ.* **2005**, *98*, 317–328.
- Dell’Acqua, F.; Gamba, P.; Ferrari, A.; Palmason, J.; Benediktsson, J.; Arnason, K. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. *IEEE Geosci. Remote Sens. Lett.* **2004**, *1*, 322–326.
- Pacifici, F.; Chini, M.; Emery, W.J. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sens. Environ.* **2009**, *113*, 1276–1292.
- Huang, X.; Zhang, L. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 257–272.
- Qin, R. An object-based hierarchical method for change detection using unmanned aerial vehicle images. *Remote Sens.* **2014**, *6*, 7911–7932.
- Qin, R.; Grün, A.; Huang, X. UAV project—Building a reality-based 3D model. *Coordinates* **2013**, *9*, 18–26.
- Mylonas, S.K.; Stavrakoudis, D.G.; Theocharis, J.B.; Mastorocostas, P. Classification of remotely sensed images using the genesis fuzzy segmentation algorithm. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5352–5376.
- Tuia, D.; Ratle, F.; Pozdnoukhov, A.; Camps-Valls, G. Multisource composite kernels for urban-image classification. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 88–92.
- Hester, D.B.; Cakir, H.I.; Nelson, S.A.; Khorram, S. Per-pixel classification of high spatial resolution satellite imagery for urban land-cover mapping. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 463–471.
- Zhang, L.; Huang, X.; Huang, B.; Li, P. A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2950–2961.
- Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *6*, 610–621.
- Shackelford, A.K.; Davis, C.H. A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1920–1932.
- Yoo, H.Y.; Lee, K.; Kwon, B.-D. Quantitative indices based on 3D discrete wavelet transform for urban complexity estimation using remotely sensed imagery. *Int. J. Remote Sens.* **2009**, *30*, 6219–6239.
- Pesaresi, M.; Benediktsson, J.A. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 309–320.
- Zhang, Y. Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. *ISPRS J. Photogramm. Remote Sens.* **1999**, *54*, 50–60.
- Huang, X.; Zhang, L. A multiscale urban complexity index based on 3D wavelet transform for spectral-spatial feature extraction and classification: An evaluation on the 8-channel worldview-2 imagery. *Int. J. Remote Sens.* **2012**, *33*, 2641–2656.
- Comaniciu, D.; Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619.
- Qin, R. A mean shift vector-based shape feature for classification of high spatial resolution remotely sensed imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 1974–1985.
- Vincent, L. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Trans. Image Process.* **1993**, *2*, 176–201.

23. Benediktsson, J.A.; Pesaresi, M.; Amason, K. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1940–1949.
24. Walter, V. Object-based classification of remote sensing data for change detection. *ISPRS J. Photogramm. Remote Sens.* **2004**, *58*, 225–238.
25. Huang, X.; Zhang, L.; Gong, W. Information fusion of aerial images and LIDAR data in urban areas: Vector-stacking, re-classification and post-processing approaches. *Int. J. Remote Sens.* **2011**, *32*, 69–84.
26. Chaabouni-Chouayakh, H.; Reinartz, P. Towards automatic 3D change detection inside urban areas by combining height and shape information. *Photogramm. Fernerkund. Geoinf.* **2011**, *2011*, 205–217.
27. Qin, R. Change detection on LOD 2 building models with very high resolution spaceborne stereo imagery. *ISPRS J. Photogramm. Remote Sens.* **2014**, *96*, 179–192.
28. Qin, R.; Gruen, A. 3D change detection at street level using mobile laser scanning point clouds and terrestrial images. *ISPRS J. Photogramm. Remote Sens.* **2014**, *90*, 23–35.
29. Qin, R.; Huang, X.; Gruen, A.; Schmitt, G. Object-based 3-D building change detection on multitemporal stereo images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *5*, 2125–2137.
30. Qin, R.; Tian, J.; Reinartz, P. Spatiotemporal inferences for use in building detection using series of very-high-resolution space-borne stereo images. *Int. J. Remote Sens.* **2015**, doi:10.1080/01431161.2015.1066527.
31. Gu, Y.; Wang, Q.; Jia, X.; Benediktsson, J.A. A novel MKL model of integrating LIDAR data and MSI for urban area classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5312–5326.
32. Cramer, M. The DGPF-test on digital airborne camera evaluation overview and test design. *Photogramm.-Fernerkund.-Geoinf.* **2010**, *2010*, 73–82.
33. Tuia, D.; Pacifici, F.; Kanevski, M.; Emery, W.J. Classification of very high spatial resolution imagery using mathematical morphology and support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3866–3879.
34. Huang, X.; Zhang, L. A multidirectional and multiscale morphological index for automatic building extraction from multispectral Geoeye-1 imagery. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 721–732.
35. Huang, X.; Zhang, L. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 161–172.
36. Qin, R.; Fang, W. A hierarchical building detection method for very high resolution remotely sensed images combined with dsm using graph cut optimization. *Photogramm. Eng. Remote Sens.* **2014**, *80*, 37–48.
37. Christoudias, C.M.; Georgescu, B.; Meer, P. Synergism in low level vision. In Proceedings of the 16th International Conference on Pattern Recognition, Quebec, ON, Canada, 11–15 August 2002; IEEE: Quebec, ON, Canada, 2002; pp. 150–155.
38. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *6*, 679–698.
39. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
40. Gruen, A.; Huang, X.; Qin, R.; Du, T.; Fang, W.; Boavida, J.; Oliveira, A. Joint processing of UAV imagery and terrestrial mobile mapping system data for very high resolution city modeling. *ISPRS J. Photogramm. Remote Sens.* **2013**, *1*, 175–182.
41. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341.
42. Wenzel, K.; Rothermel, M.; Fritsch, D. Sure—the IFP software for dense image matching. *Photogramm. Week* **2013**, *13*, 59–70.



© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).