remote sensing

MDPI

Article

Automatic Labelling and Selection of Training Samples for High-Resolution Remote Sensing Image Classification over Urban Areas

Xin Huang [1,2,*], Chunlei Weng [2], Qikai Lu [2], Tiantian Feng [3] and Liangpei Zhang [2]

Received: 22 June 2015; Accepted: 23 November 2015; Published: 1 December 2015
Academic Editors: Giles M. Foody, Norman Kerle and Prasad S. Thenkabail

1 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
2 The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; wengchunlei@126.com (C.W.); qikai_lu@hotmail.com (Q.L.); zlp62@whu.edu.cn (L.Z.)
3 College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China; Fengtiantian@tongji.edu.cn
* Correspondence: huang_whu@163.com; Tel./Fax: +86-27-6877-8525

Abstract: Supervised classification is the commonly used method for extracting ground information from images. However, for supervised classification, the selection and labelling of training samples is an expensive and time-consuming task. Recently, automatic information indexes have achieved satisfactory results for indicating different land-cover classes, which makes it possible to develop an automatic method for labelling the training samples instead of manual interpretation. In this paper, we propose a method for the automatic selection and labelling of training samples for high-resolution image classification. In this way, the initial candidate training samples can be provided by the information indexes and open-source geographical information system (GIS) data, referring to the representative land-cover classes: buildings, roads, soil, water, shadow, and vegetation. Several operations are then applied to refine the initial samples, including removing overlaps, removing borders, and semantic constraints. The proposed sampling method is evaluated on a series of high-resolution remote sensing images over urban areas, and is compared to classification with manually labeled training samples. It is found that the proposed method is able to provide and label a large number of reliable samples, and can achieve satisfactory results for different classifiers. In addition, our experiments show that active learning can further enhance the classification performance, as active learning is used to choose the most informative samples from the automatically labeled samples.

Keywords: image classification; training samples; maximum likelihood classification; support vector machine; active learning

1. Introduction

Classification is one of the most vital phases for remote sensing image interpretation, and the classification model learned from the training samples should be extended and transferred in the whole image. To date, many different pattern recognition methods have been successfully applied to remote sensing classification. Maximum likelihood classification (MLC) has proved to be robust for remote sensing images, as long as the data meet the distribution assumption (e.g., a Gaussian distribution) [1]. However, MLC does not achieve satisfactory results when the estimated distribution does not represent the actual distribution of the data [2]. In such cases, a single class may contain more than one component in the feature space, the distribution of which cannot be

described well with a single MLC. A Gaussian mixture model therefore deals with such complex situations better than simple MLC [3,4]. In order to avoid the distribution assumption, researchers have introduced non-parametric classifiers, such as the multi-layer perceptron (MLP) and support vector machine (SVM). The MLP has achieved good results in high-resolution image classification [5], pixel unmixing [6], change detection [7], and combination with other classification methods [8]. SVM has proven effective in hyperspectral image classification [9], high spatial resolution image classification [10], and multi-classifier ensemble strategies [11,12]. However, like MLC, SVM also suffers from the Hughes effect with a small-size training set [13]; thus, dimensionality reduction is important when training with limited samples. Meanwhile, the random forest (RF) classifier has received much attention in recent years, due to its robustness in high-dimensional classification problems. RF employs a bagging strategy to enhance the individual decision tree (DT), which is a weak classifier [14].

All the supervised classifiers introduced above need sufficient and efficient training samples, which are usually selected and labeled by visual inspection or field survey. However, collection of representative samples is expensive, both in terms of time and money [15]. Thus, researchers have introduced semi-supervised methods to solve the problem of insufficient sampling, by considering the unlabeled samples in an image. Meanwhile, active learning [16–22] has received increasing attention in recent years, aiming to reduce the cost of training sample selection by only labelling the most uncertain samples. Active learning has been extensively studied in the existing literature, for applications such as diverse sample selection [16], multi-view strategies [17], convergence [18], optimization of field surveying [19], image segmentation [20], and domain adaptation [21]. In general, however, most of the classification algorithms depend on manually labeled samples, even though semi-supervised learning and active learning need fewer labeled samples [15,23].

To the best of our knowledge, there have been few papers discussing machine learning methods that can automatically label training samples from remote sensing images. In this context, we propose to automatically select and label the training samples on the basis of a set of information sources, e.g., the morphological building/shadow index (MBI, MSI) [24,25], the normalized difference water/vegetation index (NDWI, NDVI) [26,27], the HSV color space, and open-source geographical information system (GIS) data, e.g., road lines from OpenStreetMap (OSM) [28]. These information indexes can be automatically obtained from remote sensing images, and hence have the potential to select and label training samples for buildings, shadow, water, vegetation, and soil, respectively. The objective of this study is to automate remote sensing image classification, and alleviate the intensity of manual image interpretation for choosing and labelling samples. Please note that active learning is a tool for selecting the most informative samples, but not for automatically labelling them. However, the proposed method can simultaneously select and label samples, especially for high-resolution images. An interesting point in this paper is that the OSM, which publically provides detailed road lines all around the world, is used to generate the samples of roads. Subsequently, in order to automatically collect reliable training samples, a series of processing steps are proposed to refine the initial samples that are directly extracted from the indexes and OSM, including removing overlaps, removing borders, and semantic constraints. In the experiments, four test datasets, as well as a large-size dataset, were used to evaluate the proposed method, using four state-of-the-art classifiers: MLC, SVM, MLP, and RF.

Section 2 introduces the information indexes, the HSV color system, and OSM, based on which the automatic sampling method is proposed in Section 3. The classifiers considered in this study are briefly described in Section 4. The experimental results are provided in Section 5, followed by a detailed analysis and discussion in Section 6. Section 7 concludes the paper.

2. Information Sources for Automatic Sample Collection

In this section, the four information indexes used to generate the initial candidate training samples are introduced. MBI, MSI, NDWI, and NDVI can automatically indicate the information

classes of building, shadow, water, and vegetation, respectively [29]. The HSV color space is used to describe the distribution of the soil and road lines are provided by OSM. In [29], these multiple information indexes were integrated and interpreted by a multi-kernel learning approach, aiming to classify high-resolution images. In addition, the MBI has proven effective for building change detection, where the change in the MBI index is considered as the condition for building change in urban areas [30].

Morphological Building Index (MBI): Considering the fact that the relatively high reflectance of roofs and the spatially adjacent shadows lead to the high local contrast of buildings, the MBI aims to build the relationship between the spectral-structural characteristics of buildings and the morphological operators [25]. It is defined as the sum of the differential morphological profiles (DMP) of the white top-hat (W-TH):

$$\text{MBI}=\sum_{d,s}\text{DMP}_{\text{W-TH}(d,s)} \tag{1}$$

$$\text{W-TH}(d,s)=\text{I-}\gamma_{\text{I}}^{re}(d,s) \tag{2}$$

where γ_{I}^{re} represents the opening-by-reconstruction of the brightness image (I), and d and s denote the parameters of direction and scale, respectively. The white top-hat DMP is used to represent the local contrast of bright structures, corresponding to the candidate building structures [25].

Morphological Shadow Index (MSI): Considering the low reflectance and the high local contrast of shadow, the MSI can be conveniently extended from the MBI by replacing the white top-hat (W-TH) with the black top-hat (B-TH):

$$\text{MSI}=\sum_{d,s}\text{DMP}_{\text{B-TH}(d,s)} \tag{3}$$

$$\text{B-TH}(d,s)=\varphi_{\text{I}}^{re}(d,s)\text{-I} \tag{4}$$

where φ_{I}^{re} represents the closing-by-reconstruction of the brightness image, and is used to represent the local contrast of shadows [25]. The MBI and the MSI have achieved satisfactory results in terms of accuracies and visual interpretation in experiments [24,25]. In this study, they are used to generate the initial training samples for buildings and shadows, respectively.

Normalized Difference Water Index (NDWI): Water has a low reflection in the infrared channel and a high reflection in the green channel [26]. Therefore, the NDWI makes use of this difference to enhance the description of water, and is defined as:

$$NDWI=\frac{B_{Green}-B_{NIR}}{B_{Green}+B_{NIR}} \tag{5}$$

Normalized Difference Vegetation Index (NDVI): According to the different reflection of vegetation canopies in the NIR and red channels [27], the NDVI is defined as:

$$NDVI=\frac{B_{NIR}-B_{\text{Red}}}{B_{NIR}+B_{\text{Red}}} \tag{6}$$

HSV Color System: HSV is a common color system, standing for hue (0~1), saturation (0~1), and value (0~1). The HSV color system is able to quantitatively describe the color space for an image [31]. In this research, HSV transform is used to detect the soil components which present as yellow or yellowish-red in the color space.

Open Street Map (OSM): OSM is a free, editable map of the whole world, which contains a large amount of location information, especially abundant and detailed road lines [28]. In this research, the road networks are registered with the corresponding remote sensing images, and the training samples for roads can then be obtained. As shown in Figure 1, accurate and reliable road sample detection consists of four steps: (a) registering the OSM road lines with the image; (b) removing the

samples labeled by other classes; (c) removing short lines; and (d) sample extraction by buffering the road lines.

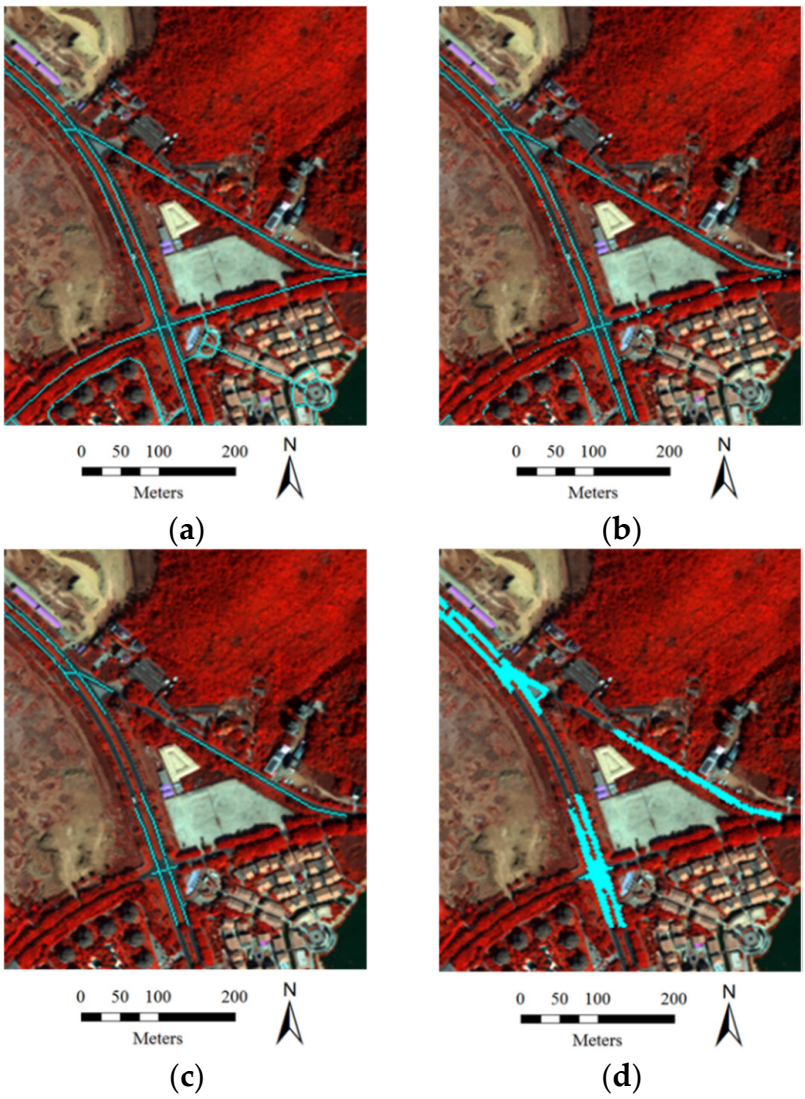


Figure 1. Detection of road samples: (**a**) registering OSM road lines with the image; (**b**) removing samples labeled by other classes; (**c**) removing short lines; and (**d**) sample extraction by buffering road lines.

A graphic example of a WorldView-2 image is used to show the effectiveness of the information indexes, the HSV-based soil detection, as well as the OSM road lines for the automatic sample collection (Figure 2). From the illustrations, it can be clearly seen that these information sources can provide effective descriptions of buildings, shadow, water, vegetation, soil, and roads. The visual results show that it is possible to automatically select candidate training samples. In particular, the soil components are highlighted as dark green in the HSV space, and can be detected as soil. However, it should be noted that there are overlaps between some similar classes, e.g., water and shadows. This suggests that the samples generated from the information sources cannot be directly used for classification, and refinement processing is needed.

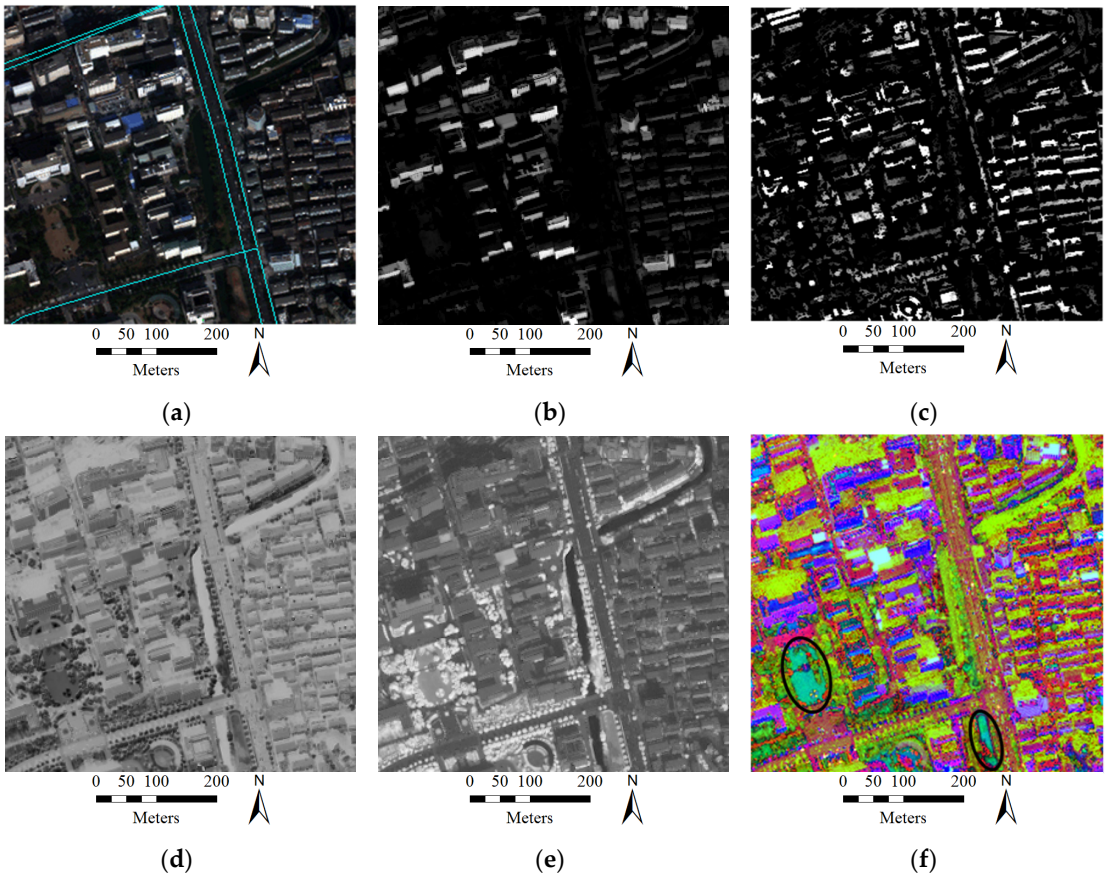


Figure 2. Multiple information sources, including (**a**) Roads from OSM; (**b**) MBI; (**c**) MSI; (**d**) NDWI; (**e**) NDVI; and (**f**) the HSV space, for selecting the initial sample set for roads, buildings, shadow, water, vegetation, and soil, respectively.

3. Automatic Sample Selection

This section introduces the proposed method for the automatic selection of training samples for buildings, shadow, water, vegetation, roads and soil, as illustrated in Figure 3, including the following steps.

(1) Select initial training samples of buildings, shadow, water, vegetation, soil, and roads, respectively, from the multiple information sources.

(2) Samples located at the border areas are likely to be mixed pixels, and, thus, it is difficult to automatically assign these pixels to a certain label. In order to avoid introducing incorrect training samples, these border samples are removed with an erosion operation.

(3) Manual sampling always prefers homogeneous areas and disregards outliers. Therefore, in this study, area thresholding is applied to the candidate samples, and the objects whose areas are smaller than a predefined value are removed.

(4) The obtained samples should be further refined, in order to guarantee the accuracy of the samples. Considering the fact that buildings and shadows are always spatially adjacent, the distance between the buildings and their neighboring shadows should be smaller than a threshold, which is used to remove unreliable buildings and shadows from the sample sets. Meanwhile, the road lines obtained from OSM are widened by several pixels, forming a series of buffer areas, where road samples can be picked out.

(5) Considering the difficulty and uncertainty in labelling samples in overlapping regions, the samples that are labeled as more than one class are removed.

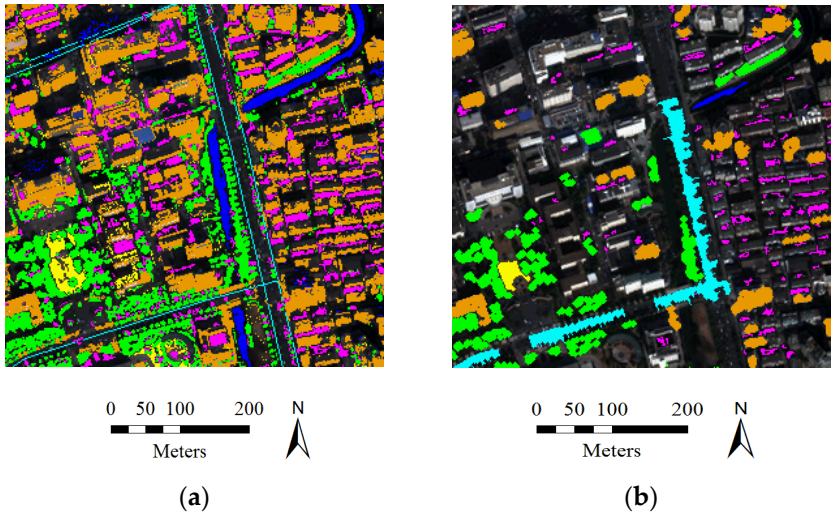


Figure 3. An example of a comparison between the initial samples derived from the multiple information sources and the samples refined by the proposed method (orange = buildings, magenta = shadow, blue = water, green = vegetation, cyan = roads, and yellow = soil). (**a**) Initial samples; (**b**) Refined samples.

The whole processing chain for the automatic sample selection is summarized in the following algorithm. Please note that the values of the parameters, mainly referring to the binarization threshold values for the multiple information indexes, and the area threshold values used to remove the small and heterogeneous points, can be conveniently determined and unified in all the test images. The suggested threshold values used in this study are not fixed, and can be appropriately tuned in different image scenes, but rather represent a first empirical approach.

Algorithm: Automatic selection of training samples

Inputs:
Multiple information sources (MBI, MSI, NDWI, NDVI, HSV, OSM).
Manually selected thresholds (T_B, T_S, T_W, T_V).
Step1: Select initial training samples from the information sources.
Step2: Erosion (SE=diamond, radius = 1) is used to remove samples from borders.
Step3: Minimal area (m^2): A_{Build} = 160, A_{Shadow} = 20, A_{Water} = 400, A_{Vege} = 200, and A_{soil} = 400.
Step4: Semantic processing:
(1) The distance between buildings and their adjacent shadows is smaller than 10 m (about five pixels in this study).
(2) Road lines are widened for buffer areas.
Step5: Remove overlapping samples.

In Figure 3, the initial samples extracted by the multiple information sources and the samples refined by the proposed algorithm are compared, from which it can be seen that the pure samples for the land-cover classes are correctly labeled in an automatic manner.

4. Classifiers Considered in This Study

In this paper, four classifiers MLC, SVM, RF, and MLP are used to implement the proposed automatic training sample selection method. The chosen classifiers have proven effective in many remote sensing applications [32].

(1) *Maximum Likelihood Classification*: MLC is a statistical approach for pattern recognition. For a given pixel, the probability of it belonging to each class is calculated, and it is assigned to the class

with the highest probability [14]. Normally, the distribution of each class in the multi-dimension space is assumed to be a Gaussian distribution. The mean and covariance matrix of MLC are obtained from the training samples, and used to effectively model the classes. If the training set is biased compared to the normal distribution, the estimated parameters will not be very accurate.

(2) *Support Vector Machine*: SVM is a binary classification method based on minimal structural risk, and it can be extended to multi-class classification with multi-class strategies. When dealing with linearly separable datasets, the optimal decision line is obtained by maximizing the margin between the two parallel hyperplanes. This type of SVM ensuring that all the samples are classified correctly is called hard-margin SVM. On the other hand, allowing the existence of misclassified training samples, soft-margin SVM introduces slack variables for each sample. SVM generates nonlinear decision boundaries by mapping the samples from a low-dimension space to a high-dimension one, and the kernel trick is used to avoid the definition of the mapping function [33].

An SVM model is constructed by support vectors, which usually locate in the decision boundary region between the class pairs. The most representative and informative samples will be close to the boundary of the class pair [15,34]. A better sample set for training an SVM model is not to accurately describe the classes, but to provide information about the decision boundary between the class pairs in the feature space [35]. Meanwhile, in the case of a small-size sample set, the outliers have an obvious influence on the decision boundary.

(3) *Neural Networks*: NNs can be viewed as a parallel computing system consisting of an extremely large number of simple processors with interconnections [32]. NNs are able to learn complex nonlinear input-output relationships, use sequential training procedures, and adapt themselves to the data [32]. The MLP is one of the most commonly used NNs. It consists of input, hidden, and output layers, where all the neurons in each layer are fully connected to the neurons in the adjacent layers. These interconnections are associated with numerical weights, which are adjusted iteratively during the training process [36]. Each hidden neuron performs a mapping of the input feature space by a transform function. After an appropriate mapping by the previous layer, the next layer can learn the classification model as a linearly separable problem in the mapped feature space, and thus NNs are able to deal with nonlinearly separable datasets [8]. In this study, the conjugate gradient method (e.g., scaled conjugate gradient, SCG), is used for training of the MLP, since it can avoid the line search at each learning iteration by using a Levenberg-Marquardt approach to scale the step size [36].

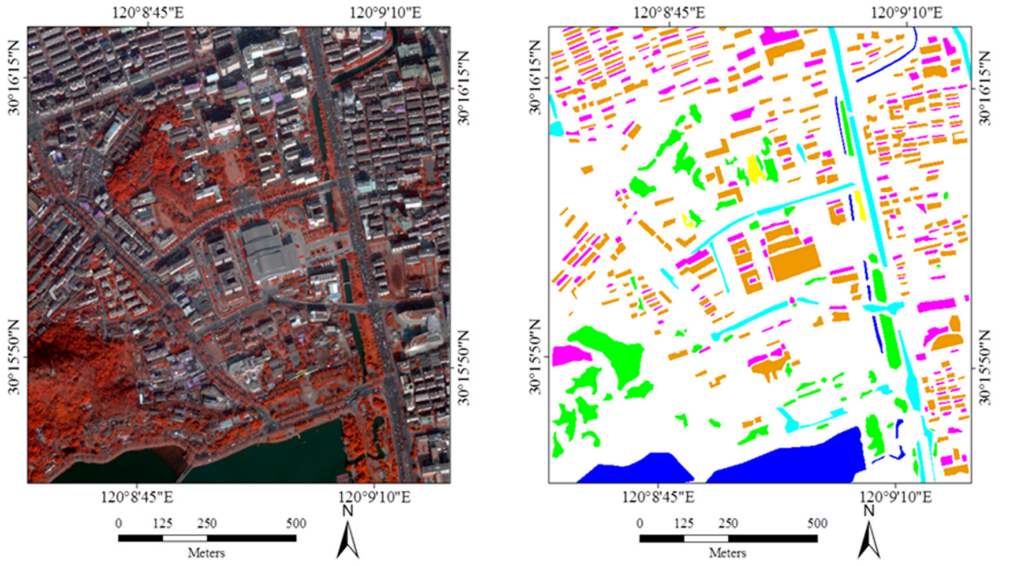
(4) *Decision Tree*: A hierarchical DT classifier is an algorithm for labelling an unknown pattern by the use of a decision sequence. The tree is conducted from the roof node to the terminal leaf, and the feature for each interior node is selected by information gain or the Gini impurity index. A pruning operation is employed in simplifying the DT without increasing errors [14]. Due to the unsatisfactory performance, an ensemble of DTs, such as RF, is more commonly used than the simple DT [37]. RF combines predictors from trees, and the final result of a sample is the most popular class among the trees. Each tree is conducted via a sub-randomly selected sample set and a sub-randomly selected feature space [38]. Each tree in RF is grown to the maximum depth, and not pruned. RF is relatively robust to outliers and noise, and it is insensitive to over-fitting [14].

5. Experiments and Results

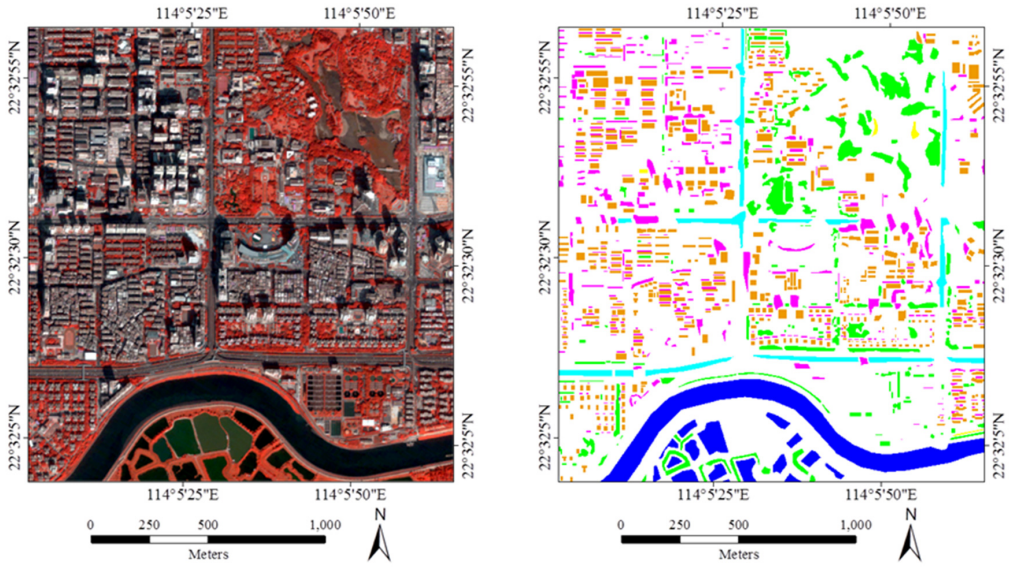
A series of test images were used to validate the proposed method for the automatic selection of training samples. In the experiments, the proposed method was compared with the traditional method (*i.e.*, manually collected samples), in order to verify the feasibility of the automatically selected samples.

5.1. Datasets and Parameters

Figure 4 shows the four test datasets, as well as the manually selected samples (the ground truth). The study areas are located in Hangzhou, Shenzhen, Hong Kong, and Hainan, respectively, with a size of 640 × 594, 818 × 770, 646 × 640, and 600 × 520 in pixels, as well as a resolution of 2 m, 2.4 m, 2 m, and 2 m. These datasets were acquired by WorldView-2, GeoEye-1, WorldView-2, and WorldView-2, respectively, with eight, four, four, and eight spectral bands, respectively The study areas exhibit the characteristics of a set of typical urban landscapes in China, and mainly consist of six classes: buildings, shadow, water, vegetation, roads, and bare soil. The six classes can be automatically extracted by the proposed method, the effectiveness of which was tested in the experiments. For the manually collected samples, 40% of the labeled samples were randomly selected from the ground truth as the training sample set (named ROI in the following text), while the rest were used for the testing (Table 1).



(**a**)



(**b**)

Figure 4. *Cont.*

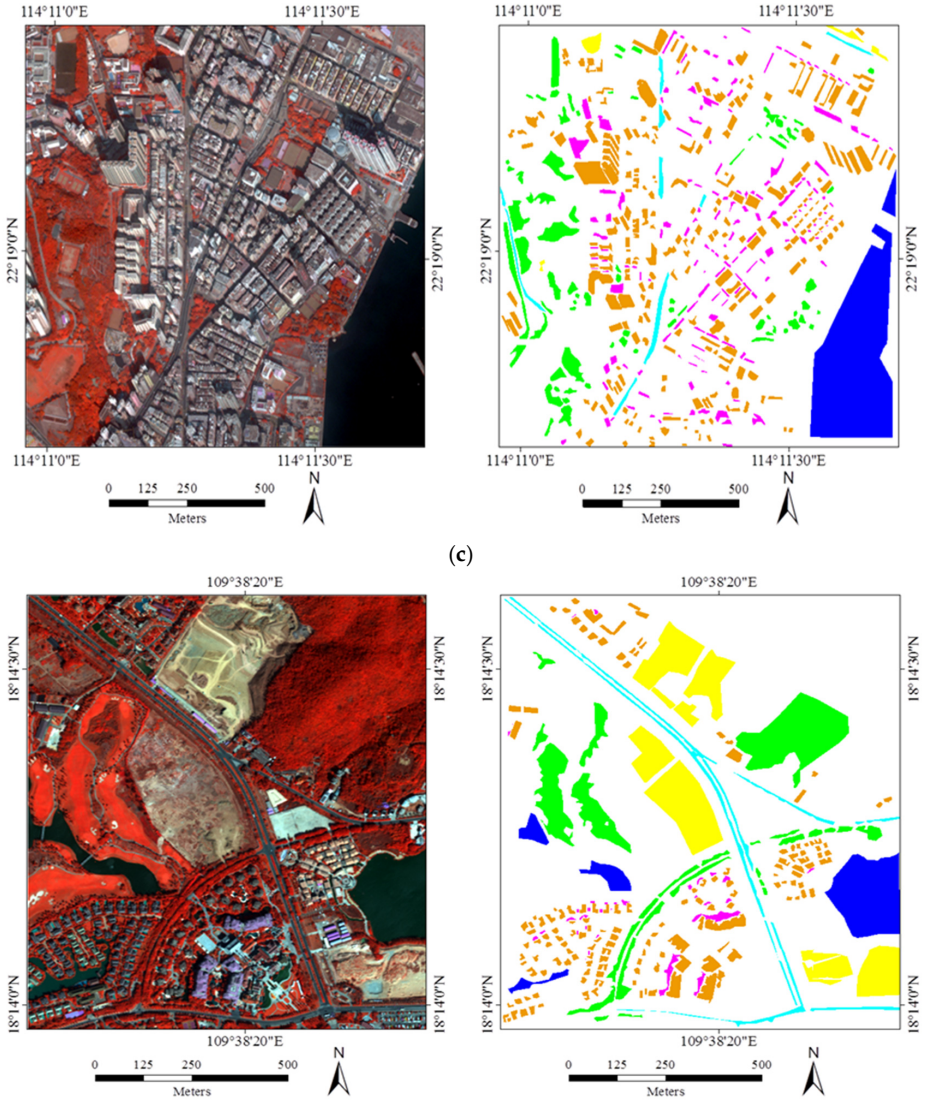


Figure 4. The images (left) and ground truth (right) used in the experiments: (**a**) Hangzhou; (**b**) Shenzhen; (**c**) Hong Kong; and (**d**) Hainan (orange = buildings, magenta = shadow, blue = water, green = vegetation, cyan = roads, and yellow = soil).

Table 1. Numbers of training samples (ROI) and test samples, which were manually generated (in pixels).

	Hangzhou		Shenzhen		Hong Kong		Hainan	
	ROI	**Test**	**ROI**	**Test**	**ROI**	**Test**	**ROI**	**Test**
Building	13,004	19,507	16,598	24,897	12,388	18,583	4631	6947
Shadow	5742	8614	9774	14,661	3788	5682	570	857
Water	6186	9279	13,759	20,639	11,725	17,588	4,483	6726
Vegetation	7118	10,678	12,829	19,244	6042	9063	8601	12,902
Road	3784	5678	5185	7778	1264	1898	2142	3214
Soil	400	601	256	385	496	746	8875	13,314

The parameters of the linear structuring element (SE) for the MBI and the MSI, including the minimal value, maximal value, and the interval, $\{S_{min}, S_{max}, \Delta S,\}$, can be determined according to the spatial size of the buildings and the spatial resolution of the images used. These parameters

were unified in this study as: S_{min} = 8 m, S_{max} = 70 m, and ΔS = 2 m, respectively. In addition, the binarization threshold values for the information indexes were set according to the suggestions of our previous study [25]. Please note that these thresholds can be simply and conveniently determined since we merely aim to choose pure and reliable samples for the further consideration.

For the SVM classifier, the radial basis function (RBF) kernel was selected, and the regularization parameter and kernel bandwidth was optimized by five-fold cross-validation. For the RF classifier, 500 trees were constructed. The MLP classifier was carried out with two hidden layers, and the number of neurons in each layer was also optimized by five-fold cross-validation.

In the experiments, training with ROI or Auto means that the classification model was trained with manually labeled samples or automatically labeled samples, respectively. Each classification was conducted 10 times with different initial training samples that were randomly chosen from the candidate training sample set, and the average accuracies were recorded as the classification accuracy. For each classification experiment, 100 training samples per class were used for the training.

5.2. Results

The automatically selected training samples of the four datasets are displayed in Figure 5, and their numbers are provided in Table 2. It can be clearly observed that the automatically labeled samples are correct, pure, and representative, and are uniformly distributed in the whole image.

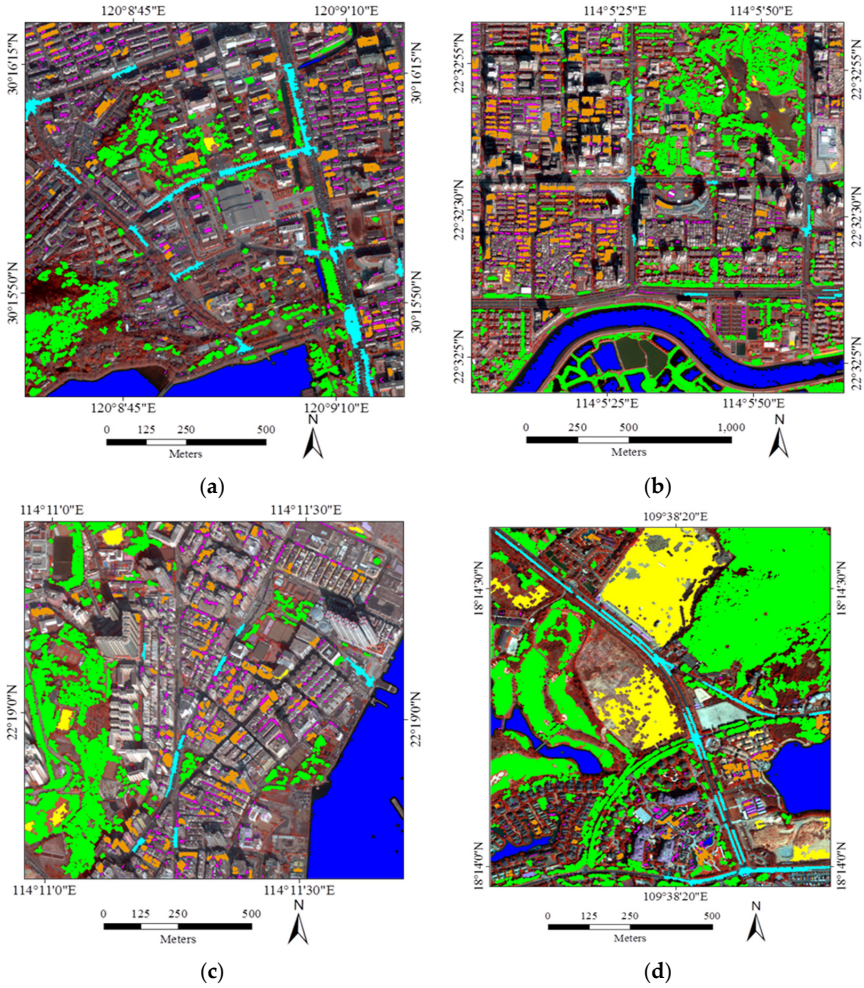


Figure 5. Automatically labeled samples from the four datasets (orange = buildings, magenta = shadow, blue = water, green = vegetation, cyan = roads, and yellow = soil). (**a**) Hangzhou; (**b**) Shenzhen; (**c**) Hong Kong; (**d**) Hainan.

Table 2. Numbers of automatically selected training samples (in pixels).

	Hangzhou	Shenzhen	Hong Kong	Hainan
Buildings	10,686	24,177	12,287	3208
Shadow	7282	10,306	7126	1419
Water	17,418	34,307	41,569	13,572
Vegetation	29,801	83,425	44,571	85,022
Roads	8210	2938	2211	6170
Soil	269	852	2263	27,369

In general, from Table 3, the Auto samples achieve satisfactory accuracies, which are close to the accuracies achieved by the ROI samples. In particular, for the Shenzhen and Hong Kong datasets, the classification results obtained by the Auto samples are very similar and comparable to the manually selected ones, for all the classifiers. With respect to the Hangzhou and Hainan datasets, the accuracy achieved by the proposed automatic sampling is also acceptable (80%~90%), although their accuracy scores are slightly lower than the ROI samples by an average of 4%~7%. Considering the fact that the proposed method is able to automatically select samples from the images, it can be stated that the method is effective, making it possible to avoid time-consuming manual sample selection.

Table 3. The overall classification accuracies for the four datasets.

		MLC (%)	SVM (%)	RF (%)	MLP (%)
Hangzhou	Auto	79.3 ± 1.7	80.4 ± 2.0	82.1 ± 1.2	82.6 ± 1.2
	ROI	83.1 ± 1.7	85.6 ± 0.8	86.8 ± 0.9	86.4 ± 1.2
Shenzhen	Auto	81.8 ± 0.7	84.4 ± 1.0	84.0 ± 1.1	83.1 ± 2.0
	ROI	82.8 ± 0.9	85.0 ± 0.9	85.1 ± 0.4	85.1 ± 0.8
Hong Kong	Auto	91.3 ± 0.6	90.2 ± 0.9	91.2 ± 0.5	91.0 ± 0.7
	ROI	92.2 ± 0.7	90.3 ± 1.6	90.4 ± 1.1	91.1 ± 0.9
Hainan	Auto	88.3 ± 0.9	86.5 ± 1.9	85.4 ± 0.8	86.1 ± 0.9
	ROI	94.1 ± 0.5	92.4 ± 0.7	90.2 ± 0.6	93.4 ± 0.5

When comparing the performances of the different classifiers with the Auto sampling, MLC achieves the highest accuracies in two test datasets (Hong Kong and Hainan). However, generally speaking, all the classifiers perform equally in the four test images, showing the robustness of the proposed automatic sampling method in different scenes.

5.3. Large-Size Image Testing

The previous experiments verified that the proposed automatic sampling strategy is able to achieve satisfactory classification results over the four urban images. We also tested the practicability of the automatic method by the use of a large-size image from the Shenzhen city center, which is China's first and most successful Special Economic Zone. The dataset was acquired on 25 March 2012, by the WorldView-2 satellite, covering 92 km^2 with a 2-m spatial resolution, consisting of eight spectral bands. As shown in Figure 6, the dataset (named WV-2 in the following text) covers the city center of Shenzhen, and three sub-regions are manually labeled as the source of the training samples (ROI). Please note that the whole image was manually labeled as the ground truth for testing, in order to guarantee the reliability of the experimental conclusions. The numbers of available samples for ROI, Auto, and test are provided in Table 4. The parameters used in this experiment were the same as the previous ones.

The classification results, including the quantitative accuracy scores and the classification maps, are shown in Table 5 and Figure 7, respectively. The experimental results convey the following observations:

- In general, the classification accuracies obtained by the automatic sampling are very promising (80%~85%), which shows that it is fully possible to automatically classify large-size remote sensing images over urban areas.
- By comparing the performances of the different classifiers, it can be seen that MLC achieves the highest accuracy for the Auto samples, while SVM and the MLP give the best results for the ROI samples.
- It is interesting to see that in the case of MLC, the automatic sampling strategy significantly outperforms the manual sampling by 8% in the overall accuracy.

MLC performs better than the other classifiers with the automatic sampling. This phenomenon can be explained by: (1) the difference in the properties between the Auto and ROI samples; and (2) the difference in the decision rules between the classifiers.

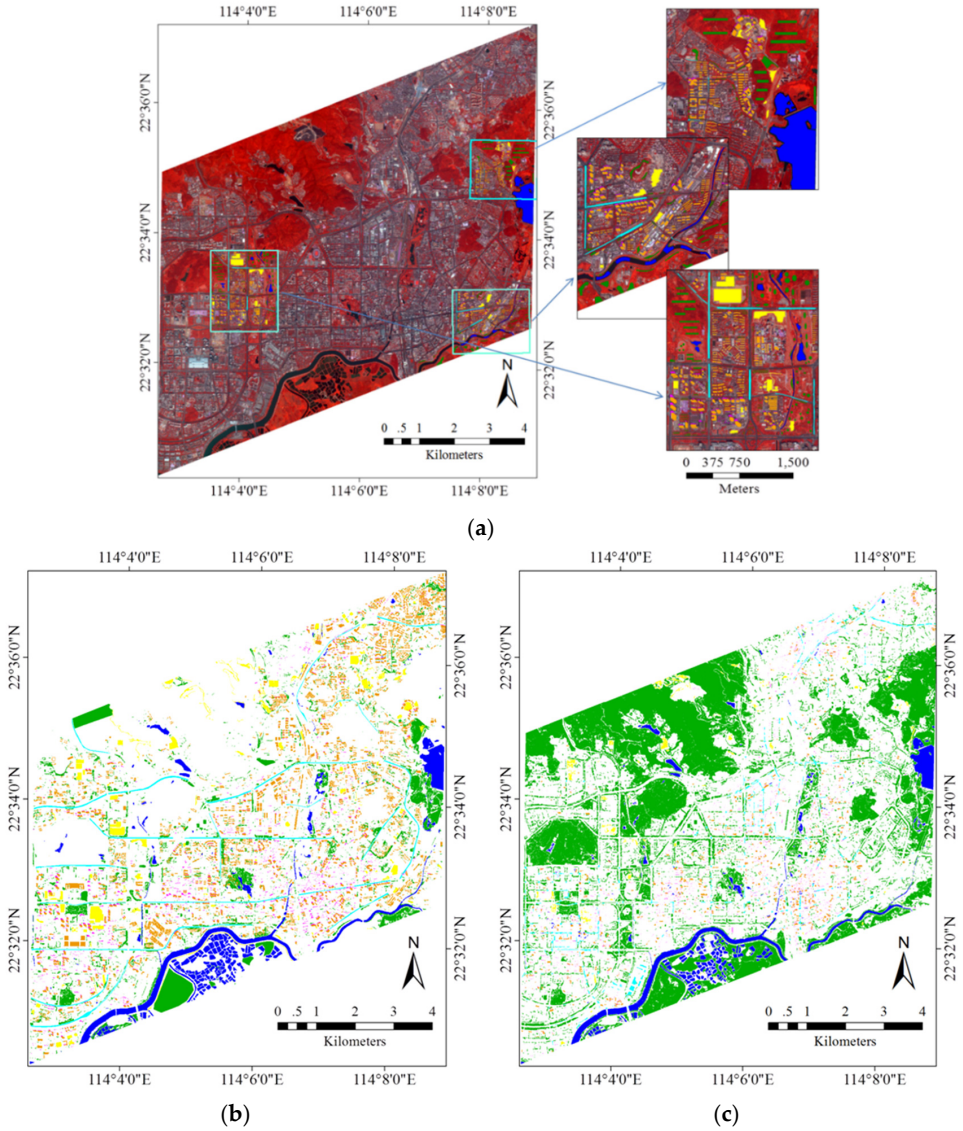


Figure 6. WV-2 Shenzhen image, as well as the three sub-regions as the source of the training sample set (ROI): ground truth of the whole image; automatically selected training sample set (Auto) (orange = buildings, magenta = shadow, blue = water, green = vegetation, cyan = roads, yellow = soil). (**a**) WV-2 urban image and training set; (**b**) Ground truth; (**c**) Automatic training set.

Table 4. Numbers of samples for ROI, Auto, and test (in pixels).

Land Cover	ROI	Test	Auto
Buildings	118,130	1,313,843	191,811
Shadow	14,446	185,650	76,723
Water	140,823	865,113	683,317
Vegetation	98,225	1,542,442	6,667,588
Roads	32,942	313,572	171,312
Soil	52,944	391,949	100,707

Table 5. The overall accuracy of the classification for the WV-2 large-size image.

Strategy	MLC (%)	SVM (%)	RF (%)	MLP (%)
Auto	84.2 ± 1.0	81.8 ± 1.7	80.9 ± 1.2	82.5 ± 0.9
ROI	76.9 ± 2.2	84.0 ± 1.5	82.6 ± 1.2	84.9 ± 1.3

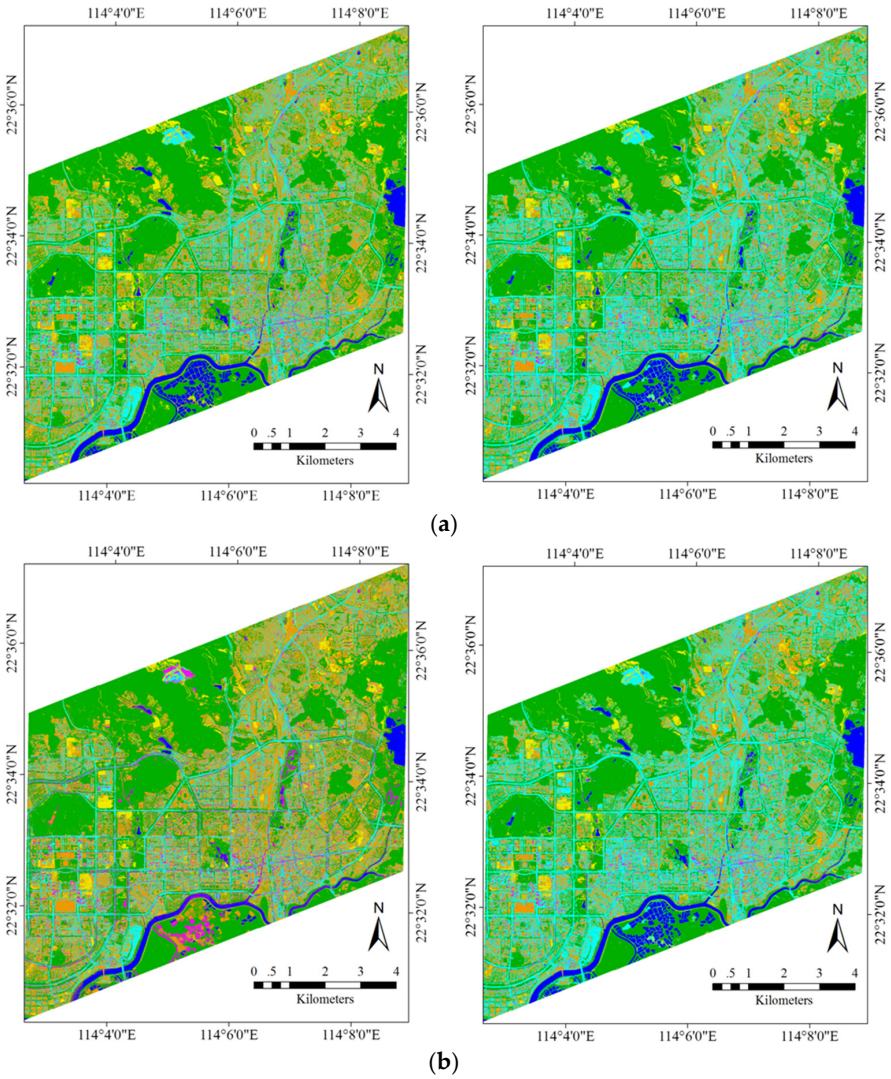


Figure 7. Classification maps of the large-size WV-2 image with the Auto and ROI training samples. (**a**) Classification maps using the Auto training samples with MLC (left) and SVM (right); (**b**) Classification maps using the ROI training samples with MLC (left) and SVM (right).

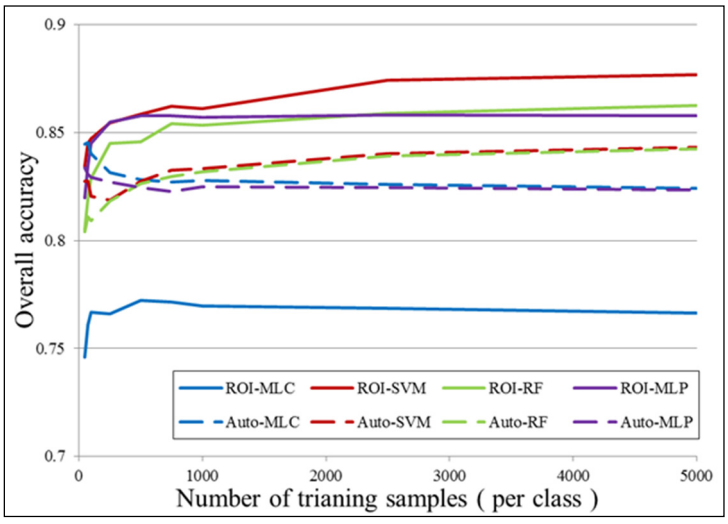
1. It should be noted that the Auto samples are purer than ROI, since the automatic selection prefers homogeneous and reliable samples in order to avoid errors and uncertainties. Specifically, as described in Algorithm 1, boundary pixels which are uncertain and mixed have been removed, and the area thresholding further reduces the isolated and heterogeneous pixels.
2. The four classifiers considered in this study can be separated into parametric classifiers (MLC), and non-parametric classifiers (SVM, RF, and MLP). The principle of MLC is to construct the distributions for different classes, but the non-parametric methods tend to define the classification decision boundaries between different land-cover classes. Consequently, pure samples are more appropriate for MLC, but an effective sampling for the non-parametric classifiers is highly reliant on the samples near the decision boundaries so that they can be used to separate the different classes.

6. Discussions

In this section, several important issues regarding the proposed automatic sampling method are discussed, including the influence of the number of training samples, the effectiveness of the classifiers for the automatic training samples, and the limitations of the proposed approach.

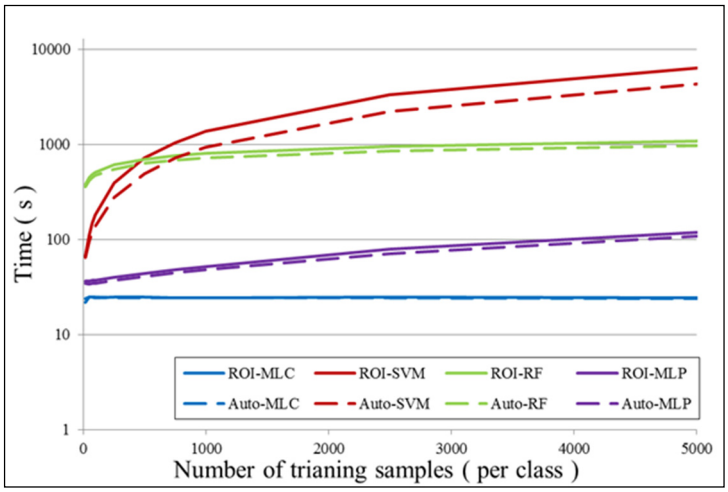
6.1. Number of Training Samples

In this experiment, the classification was conducted with different numbers of training samples extracted by ROI and Auto, respectively. The large-size WV-2 image of Shenzhen city was taken as an example. The results are demonstrated in Figure 8, where the general conclusion is that, for both ROI and Auto sampling, increasing the number of training samples does not significantly increase the classification accuracy after 500~1000 samples per class are chosen. In addition, it can be seen that Auto-MLC can provide much higher accuracies than ROI-MLC, which shows that the proposed automatic sampling method is a satisfactory sampling strategy for the MLC classifier. The CPU time for the various classifiers with different training samples is recorded in Figure 8b. Here, it can be seen that MLP and SVM are more sensitive to the number of training samples, and a large number of samples lead to more computational time. RF is less sensitive to the training sample number, as the CPU time tends to be invariant when the number of samples is larger than 1500 pixels per class. It should be noted that MLC, which aims to describe the probability of the class distribution, is totally insensitive to the number of training samples.



(**a**)

Figure 8. *Cont.*

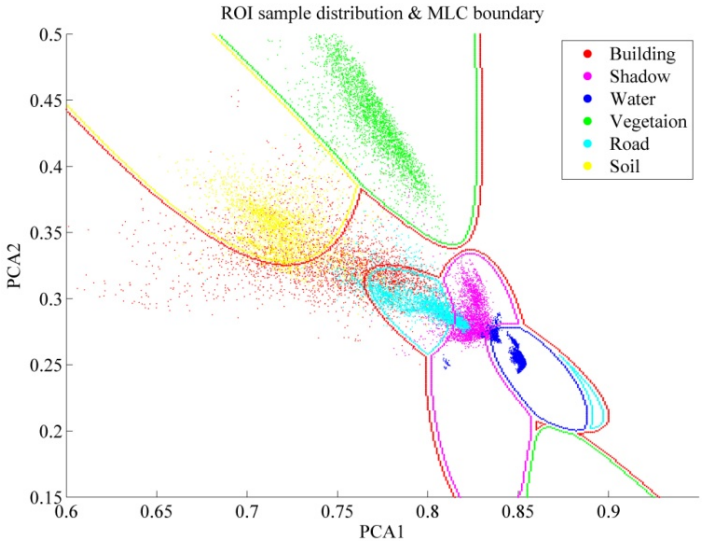


(**b**)

Figure 8. The accuracies (**a**) and computational times; (**b**) with different numbers of training samples. Auto-MLC and ROI-MLC represent the results achieved by the use of the Auto and ROI samples with MLC, respectively.

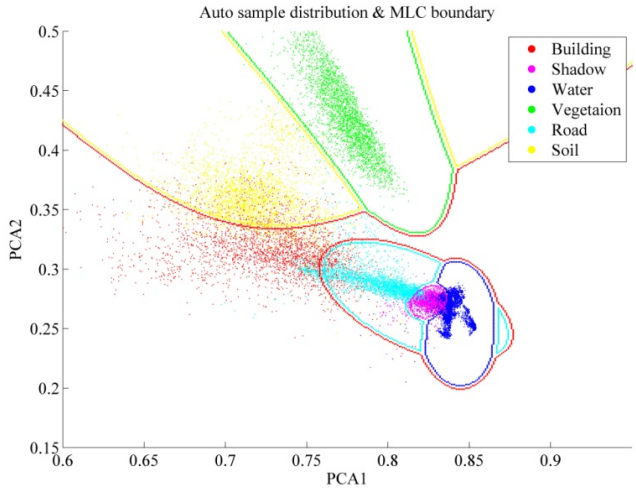
6.2. Further Comparison Between Auto and ROI Sampling

In order to further analyze and understand the automatic training samples, in this subsection, the class distributions derived from the ROI, Auto, and test samples are demonstrated and compared. In this analysis, MLC (Figure 9) and SVM (Figure 10) are taken as a representative example for the parametric and non-parametric classifiers, respectively. In the figures, the distributions are illustrated with a two-dimensional feature space (first and second principal components, namely, PCA1 and PCA2). From Figure 9, in the case of MLC, it can be clearly observed that the shape of the decision boundaries for the Auto and reference samples are quite similar. This reveals that the automatically selected samples are effective for modelling the probabilistic distributions of land-cover classes.

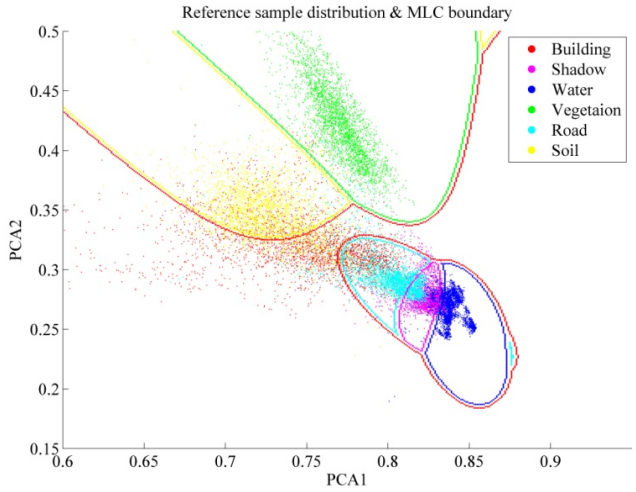


(**a**)

Figure 9. *Cont.*

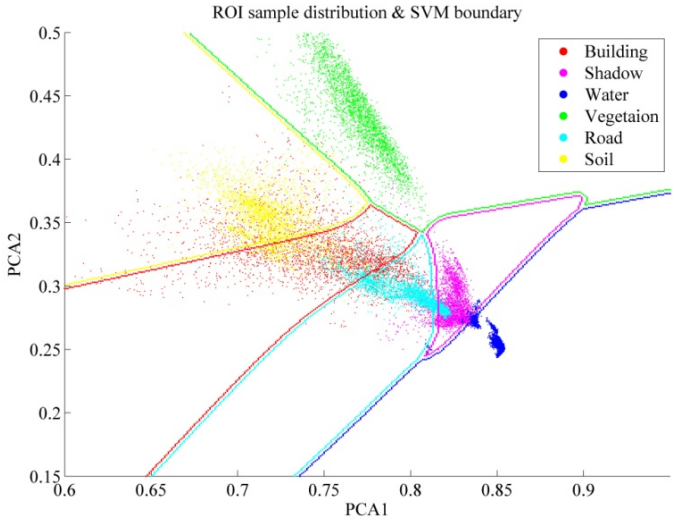


(**b**)



(**c**)

Figure 9. Decision boundaries of (**a**) ROI; (**b**) Auto; and (**c**) reference samples for the MLC classifier.



(**a**)

Figure 10. *Cont.*

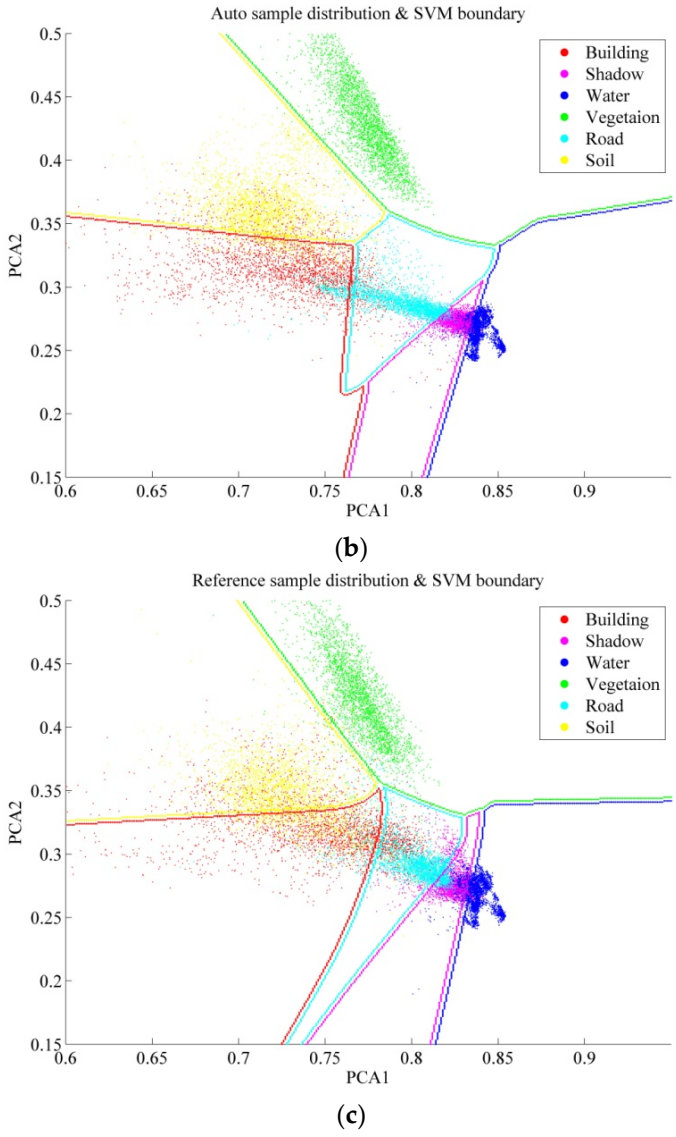


Figure 10. Decision boundaries of (**a**) ROI; (**b**) Auto; and (**c**) reference samples for the SVM classifier.

In addition, it can be seen that the decision boundaries derived from the Auto samples are more similar to those derived from the reference samples than those from the ROI samples, which can be used to explain and support the conclusion in Table 5.

On the other hand, in the case of the SVM classifier, as demonstrated in Figure 10, it can be seen that the boundaries derived from ROI are closer to the boundaries of the reference samples than the Auto sampling, which is reflected in the classification accuracy, *i.e.*, 81.8% for Auto-SVM and 84.0% for ROI-SVM, respectively. As explained previously, the accuracy of the non-parametric classifiers is highly reliant on the samples near the decision boundaries (for instance, the so-called support vectors for the SVM classification), but automatic sampling is more capable of identifying the homogeneous and pure samples which are far from the decision boundaries. Therefore, in the case of SVM, the manually labeled samples seem more suitable than the automatically selected ones. This analysis is also consistent with the accuracies reported in Table 5.

6.3. Active Learning for the Automatic Sampling

As previously shown, the proposed approach is able to automatically and effectively label samples for different land-cover classes. On the other hand, active learning can select the most

informative samples from the available sample set by considering the contribution of each sample to the classifier. In this regard, it is interesting to integrate the proposed automatic sample labelling and active learning for sample selection and optimization. The processing chain is straightforward: The candidate sample sets for various land-cover classes are selected and labeled by the proposed automatic strategy, and these samples are then ranked by active learning in terms of their contribution to the classifier. In this study, the SVM classifier is considered, and the commonly used breaking ties (BT) method [39] is adopted for the sample selection. Note that the SVM parameters (kernel parameter and penalty coefficient) are retuned during the active learning iterations. The experimental results for the active learning with the automatically labeled samples are presented in Figure 11 for the four test datasets. It can be clearly seen that in the Shenzhen, Hong Kong, and Hainan experiments, active learning (Auto-AL) can provide additional accuracy increments compared to the original automatic sampling approach (Auto-SVM). In the case of Hangzhou, active learning outperforms the original algorithm in the early stage, but does not perform as well after about 300 samples are chosen. However, in general, the difference is not significant. Therefore, it can be stated that active learning can further optimize the proposed sample labelling method, since it can select the most informative samples by considering their importance in the classification.

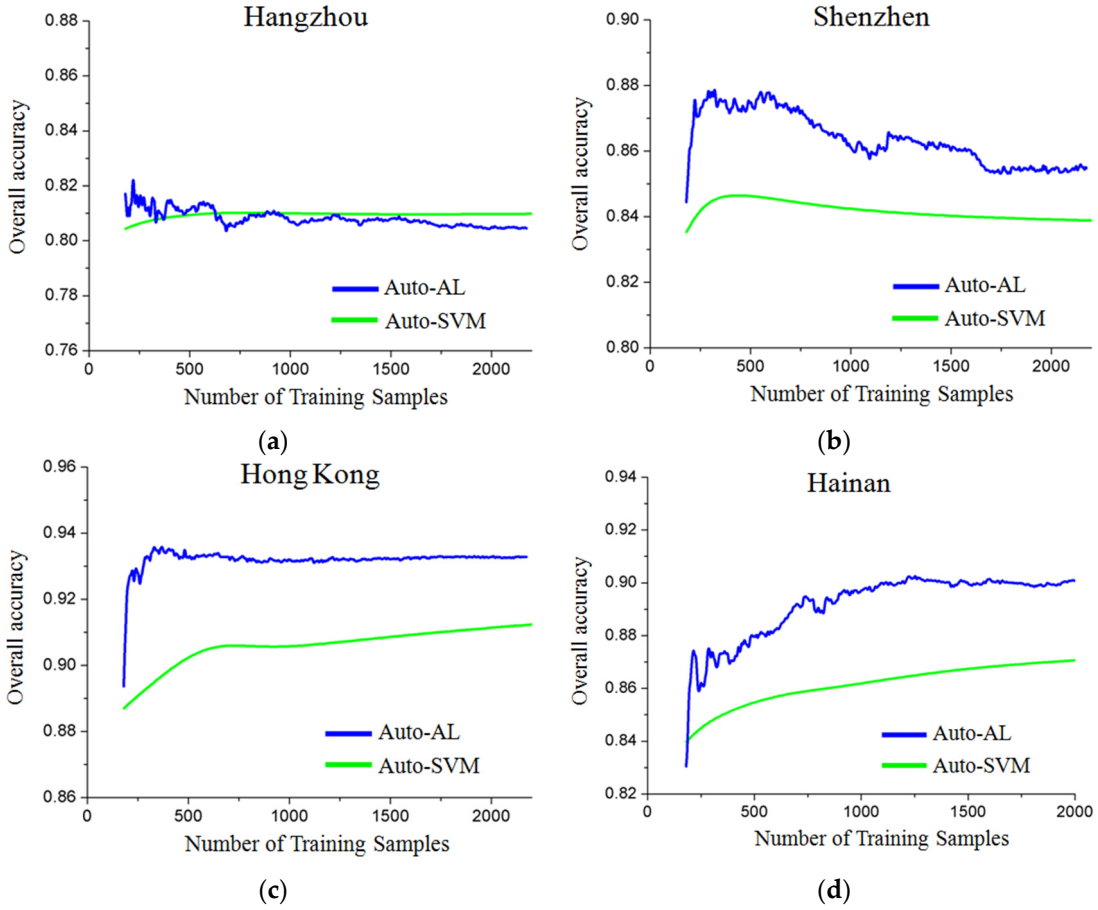


Figure 11. Accuracy curves of the active learning by using the automatically labeled samplesfor the four test images: (**a**) Hangzhou; (**b**) Shenzhen; (**c**) Hong Kong; (**d**) Hainan, respectively. Auto-AL and Auto-SVM denote SVM classification with and without active learning, respectively.

7. Conclusions and Future Scope

In this paper, a novel method for automatic sample selection and labelling for image classification in urban areas is proposed. The training sample sets are obtained from multiple

information sources, such as soil from the HSV color space, roads from OSM, and automatic information indexes referring to buildings, shadow, vegetation, and water. A series of processing steps are further used to refine the samples that are initially chosen, e.g., removing overlaps, removing borders, and semantic filtering.

The experiments with four test datasets showed that the proposed automatic training sample labelling method (Auto) is able to achieve satisfactory classification accuracies, which are very close to the results obtained by the manually selected samples (ROI), with four commonly used classifiers. Furthermore, the experiments with a large-size image (WorldView-2 image from Shenzhen city, 92 km^2) showed that the proposed method is able to achieve automatic image classification with a promising accuracy (84%). It was also found that the automatic sampling strategy is more suitable for maximum likelihood classification (MLC), which aims to describe the probabilistic distribution of each land-cover class. In particular, in the experiments, active learning [40] was jointly used with the proposed Auto sampling method, in order to select the most informative samples from the automatically labeled samples. The results were interesting and promising, as active learning could further improve the classification accuracies by about 2%~4% in most of the test sets.

The significance of this study lies in the fact that it has showed that automatic sample selection and labelling from remote sensing images is feasible and can achieve promising results. Our future research will address the mixing of manually and automatically selected samples. In this way, the decision boundaries generated by the Auto method could be further enhanced by adding new samples and removing wrong ones. It will also be possible to evaluate and compare the importance of manual and automatic samples for classification. In addition, the samples used for the accuracy assessment will be generated randomly, in order to avoid any bias in the results [41].

Acknowledgments: The authors would like to thank the editor and the anonymous reviewers for their comments and suggestions. This work was supported by the National Natural Science Foundation of China under Grant 91338111, the China National Science Fund for Excellent Young Scholars under Grant 41522110, and the Foundation for the Author of National Excellent Doctoral Dissertation of PR China (FANEDD) under Grant 201348.

Author Contributions: Xin Huang and Chunlei Weng proposed the idea of the automatic selection of training samples, undertook the experiments, and organized the writing and revision of the paper. Tiantian Feng, Qikai Lu, and Liangpei Zhang provided advice on the preparation of the experiments and the revision of the paper.

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

References

1. Wang, L.; Sousa, W.P.; Gong, P.; Biging, G.S. Comparison of IKONOS and QuickBird images for mapping mangrove species on the caribbean coast of panama. *Remote Sens. Environ.* **2004**, *91*, 432–440. [CrossRef]
2. 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. [CrossRef]
3. Dundar, M.M.; Landgrebe, D. A model-based mixture-supervised classification approach in hyperspectral data analysis. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2692–2699. [CrossRef]
4. Richards, J.A.; Kingsbury, N.G. Is there a preferred classifier for operational thematic mapping? *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2715–2725. [CrossRef]
5. Del Frate, F.; Pacifici, F.; Schiavon, G.; Solimini, C. Use of neural networks for automatic classification from high-resolution images. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 800–809. [CrossRef]
6. Licciardi, G.; Del Frate, F. A neural network approach for pixel unmixing in hyperspectral data. In Proceedings of the 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Reykjavik, Iceland, 14–16 June 2010.
7. Roy, M.; Routaray, D.; Ghosh, S.; Ghosh, A. Ensemble of multilayer perceptrons for change detection in remotely sensed images. *IEEE Geosci. Remote Sens.* **2014**, *11*, 49–53. [CrossRef]

8. Santos, A.B.; de Albuquerque Araújo, A.; Menotti, D. Combining multiple classification methods for hyperspectral data interpretation. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2013**, *6*, 1450–1459. [CrossRef]
9. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]
10. Bruzzone, L.; Carlin, L. A multilevel context-based system for classification of very high spatial resolution images. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2587–2600. [CrossRef]
11. 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. [CrossRef]
12. Huang, X.; Zhang, L. Comparison of vector stacking, multi-SVMs fuzzy output, and multi-svms voting methods for multiscale vhr urban mapping. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 261–265. [CrossRef]
13. Pal, M.; Foody, G.M. Feature selection for classification of hyperspectral data by SVM. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2297–2307. [CrossRef]
14. Mather, P.; Tso, B. *Classification Methods for Remotely Sensed Data*, 2nd ed.; Taylor and Francis: Abingdon, UK, 2004.
15. Tuia, D.; Volpi, M.; Copa, L.; Kanevski, M.; Munoz-Mari, J. A survey of active learning algorithms for supervised remote sensing image classification. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 606–617. [CrossRef]
16. Demir, B.; Persello, C.; Bruzzone, L. Batch-mode active-learning methods for the interactive classification of remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1014–1031. [CrossRef]
17. Di, W.; Crawford, M.M. View generation for multiview maximum disagreement based active learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1942–1954. [CrossRef]
18. Patra, S.; Bruzzone, L. A novel SOM-SVM-Based active learning technique for remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6899–6910. [CrossRef]
19. Persello, C.; Boularias, A.; Dalponte, M.; Gobakken, T.; Naesset, E.; Scholkopf, B. Cost-sensitive active learning with lookahead: Optimizing field surveys for remote sensing data classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6652–6664. [CrossRef]
20. Mitra, P.; Shankar, B.U.; Pal, S.K. Segmentation of multispectral remote sensing images using active support vector machines. *Pattern Recognit. Lett.* **2004**, *25*, 1067–1074. [CrossRef]
21. Persello, C.; Bruzzone, L. Active learning for domain adaptation in the supervised classification of remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4468–4483. [CrossRef]
22. Rajan, S.; Ghosh, J.; Crawford, M.M. An active learning approach to hyperspectral data classification. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1231–1242. [CrossRef]
23. Sinno Jialin, P.; Qiang, Y. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359.
24. 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. [CrossRef]
25. Huang, X.; Zhang, L. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2012**, *5*, 161–172. [CrossRef]
26. Gao, B.-C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]
27. Goward, S.N.; Markham, B.; Dye, D.G.; Dulaney, W.; Yang, J. Normalized difference vegetation index measurements from the advanced very high resolution radiometer. *Remote Sens. Environ.* **1991**, *35*, 257–277. [CrossRef]
28. Bennett, J. *Openstreetmap: Be Your Own Cartographer*; Packt Publishing: Birmingham, UK, 2010.
29. Huang, X.; Lu, Q.; Zhang, L. A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas. *ISPRS J. Photogramm. Remote Sens.* **2014**, *90*, 36–48. [CrossRef]
30. Tang, Y.; Huang, X.; Zhang, L. Fault-tolerant building change detection from urban high-resolution remote sensing imagery. *IEEE Geosci. Remote Sens Lett.* **2013**, *10*, 1060–1064. [CrossRef]
31. Wikipedia. Available online: https://en.wikipedia.org/wiki/HSL_and_HSV (accessed on 25 November 2015).
32. Jain, A.K.; Duin, R.P.W.; Mao, J. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal Mach. Intell.* **2000**, *22*, 4–37. [CrossRef]

33. Abe, S. *Support Vector Machines for Pattern Classification*; Springer: Berlin, Germany, 2010.
34. Zadrozny, B. Learning and evaluating classifiers under sample selection bias. In Proceedings of the 21th International Conference on Machine Learning, Banff, AB, Canada, 4–8 June 2004.
35. Foody, G.M.; Mathur, A. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a svm. *Remote Sens. Environ.* **2006**, *103*, 179–189. [CrossRef]
36. Møller, M.F. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw.* **1993**, *6*, 525–533. [CrossRef]
37. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
38. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [CrossRef]
39. Luo, T.; Kramer, K.; Samson, S.; Remsen, A.; Goldgof, D.B.; Hall, L.O.; Hopkins, T. In Active learning to recognize multiple types of plankton. In Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, 23–26 August 2004; Volume 473, pp. 478–481.
40. Persello, C.; Bruzzone, L. Active and semisupervised learning for the classification of remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6937–6956. [CrossRef]
41. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [CrossRef]