Article

A Multi-Primitive-Based Hierarchical Optimal Approach for Semantic Labeling of ALS Point Clouds

Xuming Ge, Bo Wu *, Yuan Li and Han Hu

Land Surveying & Geo-Informatics Department, Hong Kong Polytechnic University, Hong Kong, China; xuming.ge@polyu.edu.hk (X.G.); y-uan.li@connect.polyu.hk (Y.L.); han.hu@polyu.edu.hk (H.H.)
* Correspondence: bo.wu@polyu.edu.hk

Abstract: There are normally three main steps to carrying out the labeling of airborne laser scanning (ALS) point clouds. The first step is to use appropriate primitives to represent the scanning scenes, the second is to calculate the discriminative features of each primitive, and the third is to introduce a classifier to label the point clouds. This paper investigates multiple primitives to effectively represent scenes and exploit their geometric relationships. Relationships are graded according to the properties of related primitives. Then, based on initial labeling results, a novel, hierarchical, and optimal strategy is developed to optimize semantic labeling results. The proposed approach was tested using two sets of representative ALS point clouds, namely the Vaihingen datasets and Hong Kong's Central District dataset. The results were compared with those generated by other typical methods in previous work. Quantitative assessments for the two experimental datasets showed that the performance of the proposed approach was superior to reference methods in both datasets. The scores for correctness attained over 98% in all cases of the Vaihingen datasets and up to 96% in the Hong Kong dataset. The results reveal that our approach of labeling different classes in terms of ALS point clouds is robust and bears significance for future applications, such as 3D modeling and change detection from point clouds.

Keywords: labeling; classification; multiple primitives; ALS point clouds

1. Introduction

There are three common geometric primitives that have been used in classification issues for the past decade: points, planar segments, and voxels. Point-based classification [1] directly makes use of individual point features, as well as features dependent on neighboring points, e.g., the normal direction, surface smoothness, and local point cloud density. Planar segment- [2] and voxel-based [3] methods no longer calculate features for individual points, where the points are clustered into a primitive by using certain rules and by calculating features and assigning class labels to the primitive. Compared with the point-based method, both planar segment- and voxel-based methods show more advantages in terms of classification accuracy and speed of computation [2]. Under- and over-segmentation problems, however, are inevitable in both planar segment- and voxel-based methods. Errors arising from such problems negatively affect classification accuracy [4]. Moreover, a complex 3D scene is difficult to characterize using one type of segment. To address the challenges inherent to the characterization of a complex 3D scene that Xu et al. [4] proposed, a multiple-entity-based method is proposed, in which the authors not only combined points and planar segments but introduced mean shift segments. Different from common geometric primitives, mean shift segments are irregular in geometric space and usually derive from feature-space clustering. Xu et al.'s method [4] is a step-wise procedure. Point clouds are first classified using planar segments, after which, the points are assigned different labels. Only those points labeled "vegetation" or "unclassified" are used to carry out the second step,

namely point-based contextual classification. The mean shift segments are then extracted from the roof elements to classify these areas. Given the step-wise framework, the mean shift segments depend heavily on previous results. Moreover, only the roof elements limit the effects of mean shift segments. Taking these issues into account, we present a novel framework in the segmentation step, i.e., using multiple primitives to represent scenes. Specifically, we use the points, planar segments, and object segments in the segmentation step. The object segments in the proposed method have a clear point to the vegetation. With the introduction of the object segments, the vegetation point clouds can break the constraints of regular geometric shape and then be clustered. One of the main advantages of this segmentation strategy is that the object segments have clear meanings such that to improve the classification accuracy.

Whether or not labeled samples are used to train a classifier, semantic labeling strategies can be categorized into two groups: supervised [5] and unsupervised [6]. In this paper, we focus on the supervised strategies of ALS point clouds. Generally, a suitable classifier is trained by using labeled samples to infer testing datasets. After initial labeling for multi-class problems, an approximate optimal algorithm, such as loopy belief propagation (LBP) [7] or graph-cuts [8], can be applied based on contextual information, to smooth the labeling results. In previous work, the methods used to perform multi-class labeling and topologic optimization have been the subject of much discussion. To the best of our knowledge, most work, e.g., [2,3,5,9,10], has involved the optimization of initial labeling using an approximate optimal method in a uniform weight system. However, this procedure is problematic. Different primitives have different discriminative features with varying degrees of reliability. Thus, different weights of interaction between different primitives exist. This paper aims to refine the interaction between primitives. Specifically, when the point primitive is in a relationship, we categorize that relationship into two grades. Based on these two grades, we design a novel, hierarchical, optimal strategy to investigate the graded relationships.

In summary, the main contributions of this paper can be described thus: (1) multiple primitives are introduced to represent scenes in the segmentation step; (2) the relationships between primitives are categorized into two grades; and (3) an innovative hierarchical strategy is proposed to realize labeling optimization.

After briefly summarizing related work in Section 2, we introduce the proposed segmentation method for multiple primitives and the steps to carry out a classification in a conditional random field (CRF). In Section 3, we carry out such a classification to get the initial labeling results, which are then optimized using the proposed hierarchical method. Experiments and analysis are given in Section 4. Finally, in Section 5, we share our conclusions about the proposed method and provide an outlook on future work.

2. Related works

Point-based classification is the basic classification strategy that has attracted the majority of research [11]. Chehata et al. [1] proposed a point-based features method to classify ALS point clouds. Each point acted as an independent individual, labeled in feature space, without consideration of the relationships between other points. Niemeyer et al. [9] investigated contextual information and set up pairwise potentials in a CRF using the point primitive. Weinmann et al. [10] also used the point primitive to represent scenes and further exploited individually optimized 3D neighborhoods to calculate discriminative features. Segment-based classification is a more efficient classification strategy than a point-based strategy [2]. In the segment-based methods, points sharing homogeneous properties are grouped into the same cluster; consequently, the whole scene is split into different clusters that always contain specific meanings. One of the most important benefits is that, when compared with the point-based strategy, the features of the segment-based strategy are more discriminative, i.e., differences in some features from different types of clusters become more prominent. However, segment-based methods heavily depend on the quality of results generated by the segmentation algorithm [2,11]. In Darmawati's [12] method, the planar segments that were generated by surface growing were exploited to represent a scanning scene.

A rule-based classification method was then applied to separate buildings and vegetation in urban areas. Vosselman et al. [2] provided a combining segmentation method for classification. The authors introduced the advantages of the point-based method to overcome the disadvantages of the planar segment-based method. Voxel-based classification is another efficient classification strategy. Zhu et al. [3] proposed the generation of supervoxels to merge neighborhoods according to their features. Recently, a similar strategy using supervoxels was discussed by Lou et al. [13]. In addition to using a single primitive to represent a scanning scene, researchers also investigated how to exploit different primitives' representations. Multiple-entity-based classification [4] is considered a combination of point-based and the segment-based methods. As we described in the introduction, the authors used planar segments, points, and mean shift segments to carry out three independent classifications. Gevaert et al. [14] used spatial bins, planar segments, and local neighborhoods to represent scanning scenes in their tests. Niemeyer et al. [15] exploited two primitives (i.e., points and planar segments) in their classification method. In contrast with other work, the authors in [15] extracted planar segments based on classification results from the point-based strategy.

The derived feature vectors for each primitive are provided as an input to a classifier, so one can assign the respective semantic labels. Many studies have tried locally independent classifiers, such as the support vector machine [16], random forest (RF) [2], and AdaBoost classifier [17]. Due to its excellent performance, the RF classifier has received much attention [18]. In a recent study, many researchers [2,4,9,13,15] implemented the RF classifier to carry out the supervised classification. To obtain spatially smoothing semantic labels of 3D point clouds in multi-class cases, an approximate optimal strategy can be applied using the initial classification and contextual information [2,3]. Niemeyer et al. [5,15], Weinmann et al. [10], and Vosselman et al. [2] implemented LBP on their supervised initial classification in a CRF. From the perspective of rule-based topologic relationships, Zhu et al. [3] carried out graph-cut processing on their unsupervised initial classification in a Markov random field (MRF) to arrive at a global minimal energy. Landrieu et al. [19] proposed the integration of structures with individual points and then carried out an approximate optimal strategy from a structured regularization perspective. In the above mentioned tasks, the approximate optimal strategies were usually regarded as an independent, last step in classification.

In contrast with classic approaches, recent advances in deep learning have delivered promising results for the labeling of point clouds. Qi et al. [20] proposed PointNet to make neural networks directly applicable to 3D point clouds semantic segmentation. Recently, Landrieu and Simonovsky [21] proposed a deep learning-based framework for semantic segmentation of large-scale point clouds. In this work, the organization of 3D point clouds can be effectively captured by a super point graph that provides contextual information between object parts. Hackel et al. [22] presented a benchmark set for testing the semantic segmentation of mobile or terrestrial point clouds using the deep learning framework called Semantic3d.

3. Semantic Labeling via a Multi-Primitive-Based Hierarchical Optimal Approach

3.1. Overview of the Approach

We begin construction of the framework by separating the original point clouds into ground and non-ground point clouds [23]. The obtained ground points are exploited to generate the corresponding digital terrain model (DTM) that will be utilized to calculate the features of primitives (e.g., height). The non-ground point clouds are segmented into multi-primitives to represent scanning scenes. The RF approach is introduced to train and infer semantic labels for each unit. We then use a hierarchical approximation strategy to optimize the semantic labels in a graphical model. The overall strategy of the proposed framework is illustrated in Figure 1.
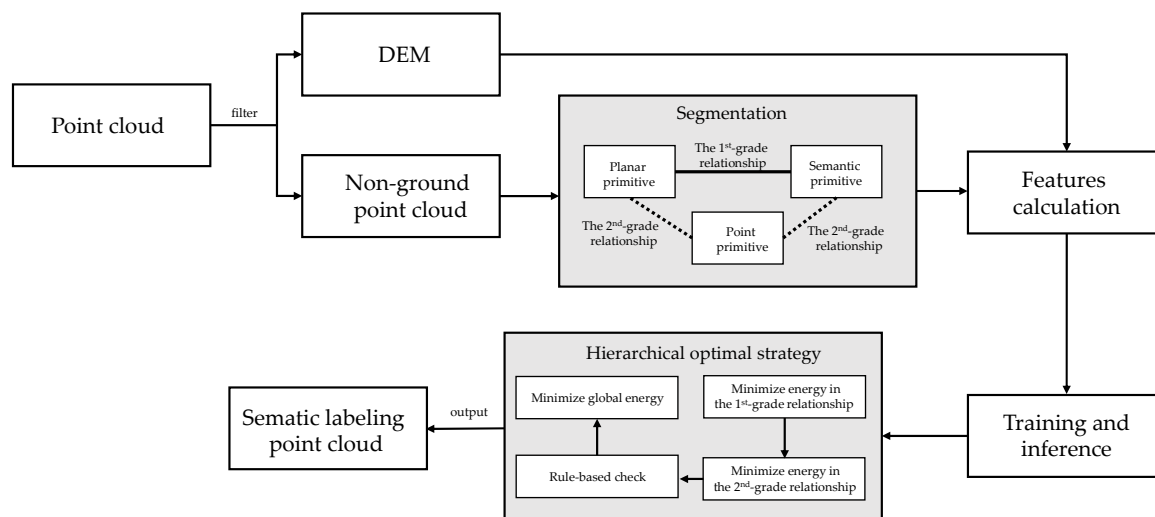
Figure 1. Framework of the proposed approach.

3.2. Multi-Primitive-Based Segmentation

We define three kinds of primitives to split and represent a scanning scene, namely, single point, planar segment, and object segment primitives. The planar segment primitive is the basic element in our strategy. The point primitive is the supplement for planar segments. Since observation errors cannot be avoided, some isolated points may not belong to any planar unit. Moreover, different objects may connect to each other and not be easily distinguished (e.g., rooves and trees). In those cases, to avoid over-segmentation, we intentionally leave those ambiguous points in primitive form, i.e., point-based. This strategy provides us an opportunity to label those ambiguous points by using the initial classification results (see Section 3.4). Otherwise, those points will be mandatorily labeled with segments. We also included an object primitive in our method to represent a scene more effectively and accurately. Considering that vegetation clusters are among the most discernible clusters in an urban point clouds scene [24], the object segments have clear points to the clusters in the proposed approach. Object segments are the expansions of planar segments (see Section 3.2.2) that make the features become more discriminative. Figure 2 shows an example of using a union set of those three primitives to represent a scanning scene. From Figure 2 we can see that if we do not keep points primitive, the pink points will be absorbed into a planar segment or be discarded as outliers. However, these operations are unreasonable because those points belong to power lines, not to roof segments or outliers. In other words, if we do not introduce object primitive, the trees (the green points in Figure 2) will be split into many small planar segments. Remarkably, that will result in a loss of many significant features of trees.
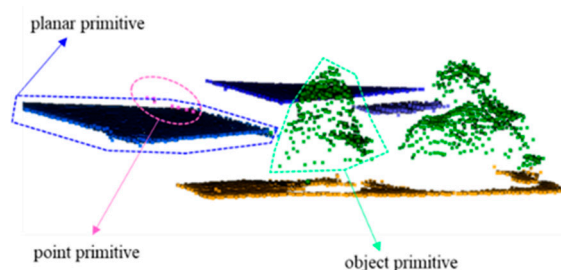

Figure 2. Multiple primitives to represent a scanning scene.

3.2.1. Planar Segmentation

The goal of planar segmentation in our method is to group points with significant planar features in a local continuous region. RANdom SAmple Consensus (RANSAC) and its improved versions [25] are often used to initiate planar segmentation. To extract planar surfaces using a RANSAC-based

method, two difficulties need to be overcome: (1) spurious planes (i.e., the planes detected by the RANSAC-based method may belong to different planes or roof surfaces) [26]; and (2) issues arising from the detection of small or narrow planes [27]. We combined a RANSAC-based method and a region-growing (RG) strategy to address these problems. Specifically, RANSAC detects planar units with a strict parameter setting scheme, to ensure the continuity of each unit. The planar units act as a seed in RG to improve segmentation, to make it as large as possible [28]. The pseudo code that shows details of this procedure is provided in Algorithm 1. In this process, the *k* nearest neighborhood algorithm is applied to detect potential seeds and to grow the region. Different constraints (e.g., local connectivity and surface smoothness) should be defined, depending on the requirements of controlling the quality of the obtained plane. In order to make the segments as large as possible and to avoid slightly non-planar surfaces breaking up into multiple, small, planar segments, we merged adjacent segments if certain constraints were first imposed. First, the normal vector directions of the two planes were nearly parallel. Second, two segments should be connected to each other. The third constraint was that a common plane be fitted with sufficient smoothness and allowable outliers (see Algorithm 1).

3.2.2. Object Segmentation

Vegetation cannot be properly captured by planar segments. Fortunately, however, certain geometrical features of vegetation are extremely different from those that can be extracted from artificial objects, e.g., buildings [24]. We therefore redefined the relationships by using both spherical and cylindrical radius searching ($r = 1$ and 0.5 m for both search modes) to set up local neighborhoods for each element. The radius setting should satisfy two issues, i.e., (1) keeping the continuity of the scanning objects and (2) avoiding the over-boundary as much as possible. Apart from satisfying those issues, we considered the related work [2,10,15] to set the searching radius. As mentioned previously, object segments are the expansions of planar segments in our approach. Thus, before executing RG, an object detection (OB) method should be executed on each planar seed (see Algorithm 1 for the exact process). The point set that includes A (A is a planar segment) and its neighborhood is marked as $\overline{A}$.

Three main constraints in our method were applied to judge the properties of $\overline{A}$, i.e., the OB module. First, the number of points in $\overline{A}$ had to be larger than that in A. Second, the variance of the normal vector directions had to be sufficiently high. Third, we projected $\overline{A}$ onto the horizontally oriented plane (noted as the XY plane) and gridded those points with a fixed step; then we could obtain subsets from $\overline{A}$ (i.e., $\overline{A} = \left\{ a_1, a_2, \cdots, a_n \mid a_i \in \overline{A} \right\}$) (see Figure 3c). For each subset, a_i with $i = 1, 2, \ldots, n$, we set up a 2D coordinate system with distance as the horizontal axis and height as the vertical axis. The leftmost/bottom point in a subset acted as the reference point to calculate the distances for other points; furthermore, the Z value of each point was the height. Figure 3 shows this process in detail. After points in one of the grid lines are reordered and connected in a designed 2D system (see Figure 3e), we can detect extreme points (e.g., the points in the red circles in Figure 3e) from the obtained curve. It clearly follows that if those points belong to a plane then the curve in Figure 3e will change to a monotonic line. In contrast, if those points belong to a tree, we will get a curve with many extreme points, such as that in Figure 3e. Therefore, a threshold of the number of extreme points (considering the noises in point clouds, we set the threshold equal to 2) can be predefined to make a judgment about whether those points belong to a plane. A contextual constraint is introduced into RG in the object segments. A point, P, has a *k* nearest neighborhood, N_p, and if all of those *k* points belong to the same object segment, then P merges with that segment. In addition, we merge adjacent object segments if they have the same intersecting boundary. The pseudo code that shows details of this procedure is also provided in Algorithm 1.

Algorithm 1. Segmentation

Input: Point Cloud = {P}.

1.　*Grid P to subset point sets SP* {sp_1, sp_2, ..., sp_n} *Calculate the normals of each sp_i and put in the corresponding Normal set $ns_i \in NS$.*
2.　***For*** *i = 0 to size (SP)*
3.　　*Obtain the planar set PS* {ps_1, ps_2, ..., ps_n} *from sp_i by a RANSAC-based algorithm.*
4.　　*Put the rest points into an outlier set OS.*
5.　　***For*** *j = 0 to size (PS)*
6.　　　*Generate a new object segment subset* oss = ∅.
7.　　　***For*** *k = 0 to size (ps_j)*
8.　　　　*Find k_n neighbors of p_k $N_{k.,}$ $p_k \in ps_j$.*
9.　　　　***If*** *size ($N_k \cup ps_j$) > size (ps_j) and oss = ∅* **then**
10.　　　　　***If*** *Do OB is true* **then**
11.　　　　　　oss ← ($N_k \cup$oss).
12.　　　　　　*Remove points from original set, i.e., SP – oss, PS - oss and OS - oss.*
13.　　　　　**End if**
14.　　　　**End if**
15.　　　　***If*** *size ($N_k \cup ps_j$) > size (ps_j) and oss ≠ ∅* **then**
16.　　　　　*Do RG to update oss, SP and OS.*
17.　　　　**End if**
18.　　　**End for** *k*
19.　　　**If** oss = ∅
20.　　　　*Do planar RG to update ps_j*, SP and OS.
21.　　　**Else**
22.　　　　*Put oss into the object segment set OSS.*
23.　　　**End if**
24.　　**End for** *j*
25.　**End for** *i*
26.　***For*** *i = 0 to size (PS)-1*
27.　　***If*** *$ps_i \neq$ ∅* **then**
28.　　　*If merge conditions are true: ps_i ← $ps_i \cup ps_j$, with j = i + 1 to size (SP), ∅ ← ps_j.*
29.　***For*** *i = 0 to size (OSS)-1*
30.　　***If*** *$oss_i \neq$ ∅* **then**
31.　　　*If merge conditions are true: oss_i ← $oss_i \cup oss_j$, with j = i + 1 to size (OSS), ∅ ← oss_j.*
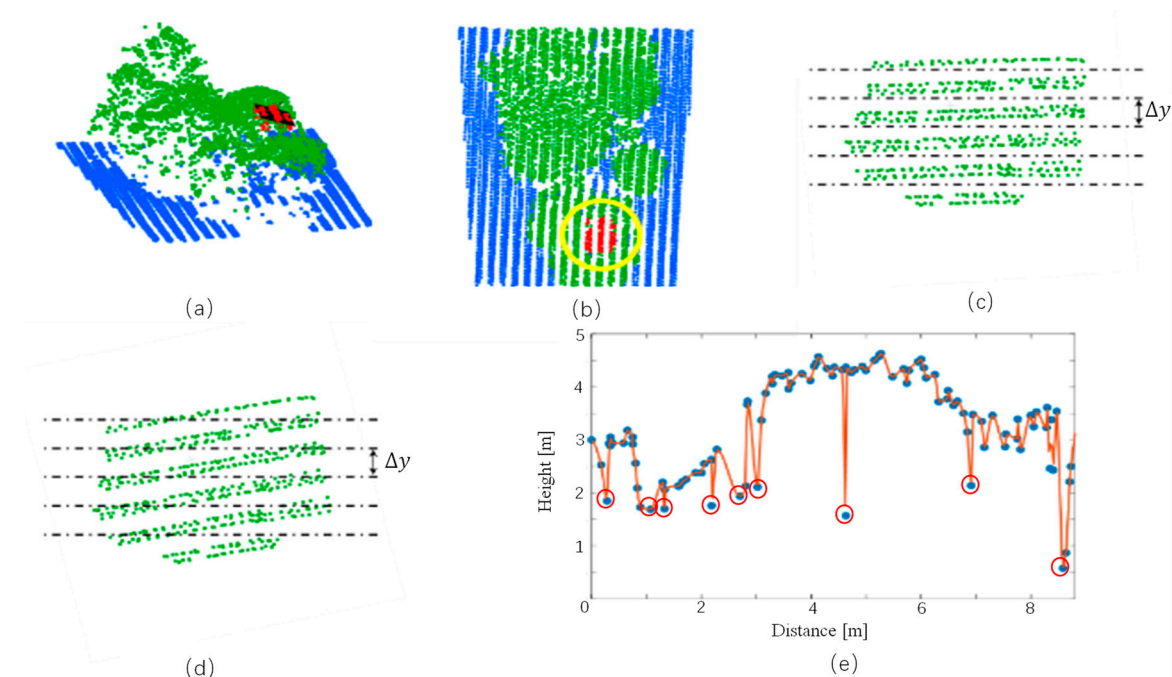32.　Output: *PS, OSS, OS.*

Figure 3. The strategy used in the proposed method to carry out a semantic check. (**a**) Point clouds in 3D space. The green and blue points represent tree and ground, respectively. The red points represent a candidate of semantic segments, e.g., plane segment A. (**b**) Point clouds in 2D space. The yellow circle shows the neighborhood using a cylinder radius. (**c**) Point clouds of $\overline{A}$ and a designed grid with a fixed step, Δy. (**d**) The scanning line is not required to be parallel to the grid. (**e**) Points in one of the grid lines are reordered in a designed 2D system and extreme points (e.g., points in the red circles) can be detected from the 2D curve.

3.3. Feature Description for Semantic Labeling

3.3.1. Unary Features

Unary features for different primitives have been much discussed in previous work [2,3,10], so here we only summarize those features pertaining to our work. All of the quantified features are normalized into the interval 0 to 1 by the min-max normalization. A detailed description is found in Table 1.

Table 1. Unary features description.

No.	Domain	Feature's Description
1	N	The number of points in segment *i*.
2	N	The perimeter and area of segment *i* in 2D space.
3	N	Feature derived from the height (calculated from spherical and cylindrical neighborhoods): average height above DTM; the difference between the maximum and minimum heights; the variance in height of segment *i*.
4	N	Features derived from the covariance matrix by principal component analysis, i.e., linearity, planarity, scattering, omnivariance, anisotropy, eigenentropy, and change of curvature.
5	N	The variances of planarity, omnivariance, and change of curvature within a cylindrical neighborhood are calculated for the point primitive only.
6	N	The number of extreme points for a cylindrical neighborhood of segment *i*.
7	[0, 360]°	Features derived from the normal vector direction.
8	N	The point densities in spherical and cylindrical neighborhoods.
9	[0, 180]°	The difference of normals of segment *i* using a multi-scale operator [21].

3.3.2. Pairwise Features

Contextual information has played a very important role in the recent classification of point clouds [2,3,13,29]. According to the different discriminative powers of different primitives, we categorize the geometric relationships into two grades. The first grade includes relationships in planar-to-planar, object-to-object, and planar-to-object segments. The second grade represents the relationships in point-to-point, point-to-planar, and point-to-object segments. Figure 4 shows the proposed two grade relationships between multiple primitives. One of the main advantages of this definition is that we can control the magnitude and direction of action of each primitive. We will discuss the advantages of this definition in detail in Section 3.4. Here, we give pairwise features in Table 2, after which, we carry out the same normalization operation as that in the calculation of unary features.
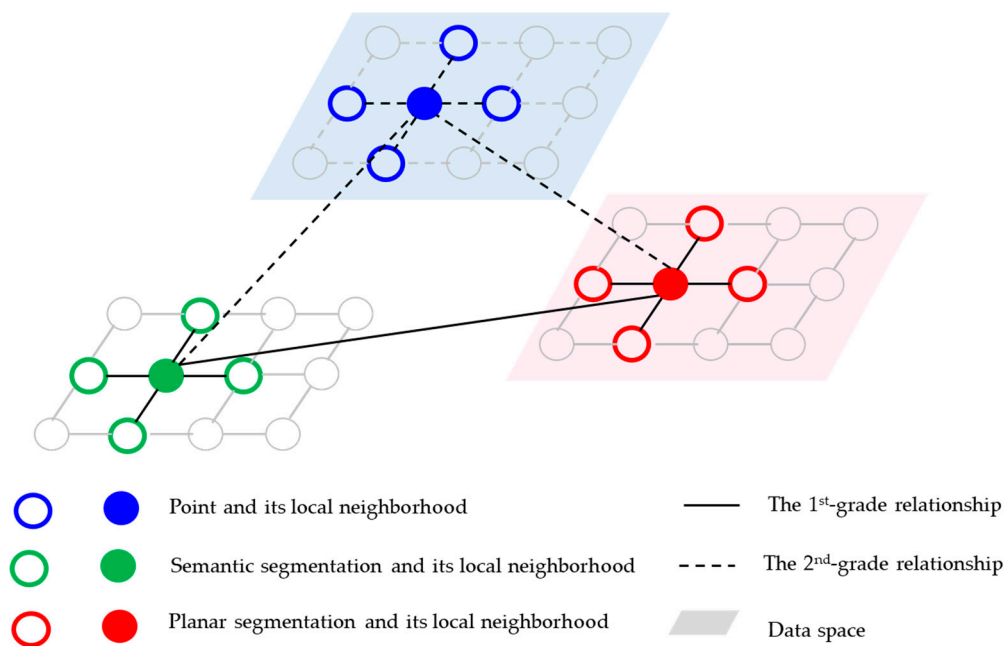

Figure 4. Relationship definition of multi-primitives in the proposed method.

Table 2. Pairwise features description.

No.	Domain	Feature's Description
1	N	The number of the 2D border intersecting points.
2	[0, 1]	The linearity of the 2D border intersecting lines.
3	N	The average height difference of adjacent segments in the 2D overlapping part.
4	[0, 360]°	The variance of the direction of the 2D border intersecting points in each segment.
5	N	The feature derives from point densities that are separately calculated using spherical and cylindrical neighborhoods.
6	{True, False}	For the point-to-point case, we define points i and j as connecting to each other if point j is in the neighborhood of point i and vice versa. The number of border points is a constant (i.e., = 1).

3.3.3. Heuristic Rule-Based Features

To emphasize the advantages gained from contextual information, researchers in recent years have paid more attention to state-of-the-art features, namely high-order features. Although high-order potentials are becoming ever more important, it is still difficult to apply them to the classification of point clouds because of their complex relationships and extensive computational costs. So far, there is no unified definition of high-order features [2]. In many applications, the regional context (namely the long-range relationships and label cost) is used to calculate high-order potentials [13,30,31].

Although such a definition makes sense for the classification of point clouds if the unary potentials are based on the point primitive or the (super) voxel primitive, it is not the best definition in cases concerning the segment primitive. Segments are already grouped to objects or to larger object parts based on the similar features so that neighboring segments are very likely to have different labels. Many researchers [3,4,13] have cited the use of heuristic rule-based knowledge acting as high-order features to correct for unfavorable labels in labeling results. Although rule-based knowledge comes from subjective observations, it has clear and correct physical meanings that heavily depend on objective factors. Therefore, such knowledge is highly reliable. Considering the high repeatability of the method, only a limited number of rule-based features is employed, see Table 3.

Table 3. Heuristic rule-based features description.

No.	Domain	Feature's Description
1	{True, False}	The area and minimal height of a roof segment should be larger than given thresholds (e.g., 15 m^2 and 1.5 m, respectively).
2	{True, False}	The edges of a roof segment should be larger than a given threshold (e.g., 1 m).
3	{True, False}	There is no isolated vegetation point in a roof segment.
4	{True, False}	A façade segment is always under a roof segment, and the angle of the normal vectors of such two segments should not be parallel.

3.4. Hierarchical Labeling Strategy

3.4.1. Conditional Random Fields (CRFs)

CRFs belong to undirected graphical models that provide a probabilistic framework for context-based classification. We apply a CRF for the proposed classification. A CRF is defined on an underlying graph $G = (n, e)$ consisting of a set of nodes, n, and a set of edges, e. Each segment or point, in our case, corresponds to a node, n_i ($n_i \in n$), and the goal is to assign an appropriate semantic label, y_k ($y_k \in Y$), for n_i with corresponding observations, x_i ($x_i \in x$). Each edge, e_{ij} ($e_{ij} \in e$), links nodes, n_i and n_j, with a certain defined neighborhood. In our case, e_{ij} represents the contextual relations between two such nodes and can then be used to support the potentials. The semantic labels for all nodes based on their own observations are predicted and collected in a vector $Y = \{y_1, y_2, \ldots, y_n\}$, where n is the number of nodes. The energy function to express the goal function in a CRF is:

$$E(Y) = \sum_{i \in n} E_u(x_i, y_k) + \alpha \sum_{e_{ij} \in e} E_p(x_{ij}, y_k, y_h) + \beta \sum_{i \in n^2} E_h(y_k, l_i). \tag{1}$$

In Equation (1), E(Y) represents global energy. Our goal is to find an appropriate combination of Y so as to minimize global energy. The terms E_u and E_p are the unary term and pairwise term, respectively. The high-order term, E_h, is designed using the label cost term introduced in [32]. l_i represents an optimal neighborhood of node, n_i. We consider the label cost in the point primitive (i.e., in Equation (1)). n^2 is a subset of nodes that only includes the point primitive, $n^2 \in n$. α and β are the corresponding weight factors.

3.4.2. Calculation of Local Energies

We used RFs to carry out semantic labeling in order to obtain initial classification results. Based on RF results, we can extract the energies from the unary and pairwise potentials, with all possible semantic labeling and their combinations. Three independent RFs, one for the unary potential and two for the pairwise potential, must be trained. Specifically, in the unary potential, the RF will be assigned to the points, planar segments, and object segments. The second RF is assigned to the relationship in the segmentations. The last RF belongs to the relationship between the segmentations and points. In this study, these classifiers are trained independently from each other with fully labeled data. The labeled point cloud is segmented by the algorithm described in Section 3.2. After segmentation, we calculate

features as described in Section 3.3 for all nodes and edges. A feature matrix is formulated as input data for each RF. To avoid an unbalanced training sample, we apply downsampling or oversampling techniques to ensure that the RF can randomly select the same number of samples for each class [33]. As mentioned in Section 3.4.1., we used a CRF for classification. The goal of classification is to find the most probable configuration labels given the observed point clouds. Thus, to arrive at that goal in a CRF, i.e., denoting the combination of all input data by x, we want to determine the configuration of class labels that maximizes the posterior probability $P(Y|x)$ [2]. Considering that in Equation (1) we used the energy function to realize the goal optimization, to keep the form consistent, the posterior probability can be transformed to energy consumption, i.e., $1 - P(Y|x)$. For the inference, we use the number of votes, N_{y_i}, to express the posterior probability of a node having a label y_k, then the corresponding unary energy function is:

$$E_u(x_i, y_k) = 1 - \exp\left(N_{y_k}/n_T\right), \tag{2}$$

where n_T is the total number of trees to be used in an RF.

For pairwise issues, we introduce only two classes to judge whether linking nodes have the same labels. Obviously, this designation will markedly reduce the number of classes in pairwise cases so that accuracy of the classification of RFs will be improved. Moreover, the required number of samples for training is also reduced considerably. Another advantage for this definition is that the functioning range of contextual information becomes clearer and more reasonable, i.e., pairwise potentials can only determine the relations of two nodes that are linked by a certain edge but cannot determine a specific label for a node.

For high-order issues, the label cost term tends to reduce redundant label categories by imposing the cost of these labels on a category subset. As mentioned in Section 3.4.1., we consider the label cost in the point primitive such that it requires fewer category labels to describe a region. The label term penalizes category y_k heavily when there are a few points labeled as y_k in a neighborhood. Moreover, penalization also acts on each neighborhood, e.g., neighborhood A is penalized more heavily than neighborhood B if A has more categories. These strategies are supported by the continuity property of an object's surface. The proposed approach to calculate the label cost values is based on [13].

3.4.3. Hierarchical Strategy

As previously mentioned, contextual information should be used to affect a semantic label determined for each 3D point. So far, exact optimization is computationally intractable in multi-class problems; hence, approximate algorithms are applied instead. We apply an energy minimum method [32] to optimize the energy function, i.e., Equation (1) within an appropriate weight factors combination.

Taking into account that the point primitive has a high possibility of sharing the same label with its neighboring segments, we introduce a penalties process with the robust PN-Potts model to smooth the labeling results piecewise:

$$E\big(y_i, y_j\big) = \varpi_{ij}\Delta\vartheta\big(\big[y_i, y_j\big]\big), \tag{3}$$

where ϖ_{ij} are the weights and $\vartheta(\Delta)$ is:

$$\vartheta\big(\big[y_i, y_j\big]\big) = \begin{cases} \min\big(\varepsilon_1, E_u\cdot\frac{\varepsilon_1 - \varepsilon_2}{\gamma} + \varepsilon_2\big), & \text{if } E_u < E_e \\ 0, & \text{else} \end{cases} \tag{4}$$

with E_u and E_e representing the energy of $y_i \neq y_j$ and $y_i = y_j$, respectively, ε_1 is an upper bound on the potentials, and ε_2 and γ are the remaining parameters of a truncated linear cost function that ensures the desired, gradual increase in the penalty. We specifically define a weight as follows: a node, i, with

m neighboring nodes that can be linked with m edges, e_{ij} (j = 1, 2, ... , m); then, we give a weight for the edge, e_{ij}, that corresponds to node i:

$$\omega_{i \leftarrow j} = \frac{1}{m} \cdot \left\{ \psi_{ij}\left(x_{ij}, y_i, y_j\right) \middle| y_i = y_j \right\}. \tag{5}$$

We carry out a hierarchical strategy to complete the optimization process. First, we implement the energy minimum method in a sub-energy function:

$$E(Y)^1 = \sum_{i \in n^1} E_u(x_i, y_k) + \alpha^1 \sum_{e_{ij} \in e^1} E_p^1\left(x_{ij}, y_k, y_h\right), \tag{6}$$

where $E(Y)^1$ is a sub-energy of $E(Y)$. E_p^1 is the pairwise potential that considers the relationships in the segmentations, i.e., the 1st grade relationship. n^1 is a subset of nodes that includes the planar and object segments and e^1 is the corresponding edges set. α^1 is a factor that controls the weight of pairwise potentials. The features in the planar and object segments have strong discriminative power, hence we can use them to calculate highly accurate semantic labeling results from the unary potential. Thus, the unary potentials should predominate in the sub-energy function, i.e., Equation (6). For α^1 we choose from [0, 1]. Based on the semantic labeling results from the first hierarchy, we implement an advanced calculation. In this calculation, the energy minimum method can be applied to the following sub-energy function:

$$E(Y)^2 = \delta \sum_{i \in n^1} E_u(x_i, y_k) + \sum_{i \in n^2} E_u(x_i, y_k) + \alpha^2 \sum_{e_{ij} \in e^2} E_p^2\left(x_{ij}, y_k, y_h\right) + \beta \sum_{i \in n^2} E_c(x_i, l_i), \tag{7}$$

where e^2 is a sub-edge set that includes the edges in the point-to-point and point-to-segmentation cases, i.e., the 2nd grade relationship. α^2 is a weight to control the relationship between the point unary potentials and the points related to pairwise potentials. δ is a factor to relate the results from the first step. β is a weight factor to control the contribution from high-order issues. It is clear that the planar and object segments promote a higher accuracy than a single point does, suggesting that contextual information from a planar or object segment greatly affects the linking point; on the contrary, the effect becomes negligible.

4. Experimental Analysis

To verify the performance of the proposed method, two test datasets with different characteristics are exploited in this paper. The first uses the ISPRS benchmark ALS datasets from Vaihingen, Germany [34]. The dataset covers a typical town with buildings of several floors. The second dataset was collected by an ALS scanning survey in the Central District of Hong Kong, which is a typical metropolitan area with many skyscrapers. The point clouds are already manually labeled by the data provider. The details of those two datasets and evaluations are described below.

4.1. Vaihingen Dataset

The Vaihingen dataset [35] for the testing of urban classification and 3D building reconstruction [34] is composed of three sites with different scenes (Figure 5). Based on the description from the ISPRS, the 3D points were manually labeled to enable an evaluation of the semantic labeling results. Generally, the Vaihingen sites have relatively low buildings, mostly lower than 20 m, with classical architecture. Area 1 (including 153,762 points) is the inner city, with dense, complex buildings, and some trees. Area 2 (including 267,189 points) consists of many high-rise, residential buildings surrounded by trees. Area 3 (including 227,517 points) is a residential zone with many small, detached houses.

Figure 5. Test sites of scene Vaihingen. From left to right: areas 1, 2, and 3 [34].

In our tests we focused on the urban regions to restrict ourselves to five class labels: *ground*, *building roof*, *façade*, *vegetation*, and *other*—the main and clear objects in the ALS point clouds. Moreover, one of the main purposes for us to label point clouds is to help us detect and extract buildings from point clouds. Thus, the training samples (including 234,000 points) were manually transformed into those five classes. The ground points were filtered before classification. However, the introduced methods could not extract all the ground points from a complex scene; consequently, an independent label was applied to find the remaining ground points.

We first filtered out the ground points from the point clouds and then implemented the proposed segment method to split the point clouds into the expected primitives. Figure 6 shows some of the segment results of the Vaihingen training dataset. Figure 6a displays a scanning scene that was split into planar and semantic segments. In Figure 6b we added the point primitive based on Figure 6a. First, we can see that the main planar surfaces and vegetation can be well segmented. Moreover, those segments are grouped to be as large as possible—-especially the semantic segments that can find vegetation point clouds with high accuracy and can grow as large as possible. In Figure 6b there are some isolated points in the scanning scene (the white points), with some clearly on top of the roof. It makes sense to separate those points from roof segments, because some of them belong to power lines and others are difficult to classify (e.g., clusters). Some isolated points appearing in a vegetation region may be caused by the variant point densities but they maintain, nonetheless, a high probability of being classified within the vegetation class by their contextual information. Other kinds of isolated points belong to a case in which a fitting plane is so small (e.g., part of a car or part of a fence and hedge) that one cannot fulfill the required qualities of a segment using our approach.
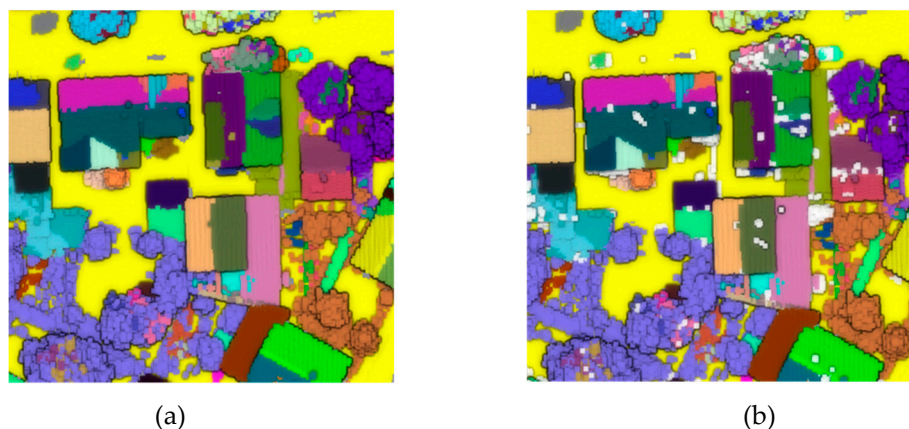

(a)   (b)

Figure 6. Multiple primitives to split point clouds. (**a**) Planar segments and semantic segments. (**b**) Segments and points.

Figure 7 shows the semantic labeling results for the three Vaihingen study areas. 3D points were labeled with predefined classes: *ground* (orange), *building roof* (blue), *vegetation* (green), *façade* (yellow), and *other* (red). First, in line with the main purpose of classification using our method, building structures satisfying the predefinition of a building area larger than 50 m^2 in our tests can be safely and effectively detected and extracted, with point clouds of reasonable quality. The minute planes were properly assigned to the *other* group, because they neither satisfied the area threshold of a building roof (15 m^2) nor resonated with the discriminative features of other classes. The façade of a building could be reasonably detected by the proposed method in such areas. From Figure 7 we can see that, although there are some mistakes in the *vegetation* class, the main vegetation areas (>5 m^2) can be detected and extracted reasonably well. The fault classification of *vegetation* is mainly in the case of the fence and hedge. Because some low vegetation twines on those objects, and we did not predefine a certain class for the fence and hedge, some parts of the objects will be divided into *vegetation* and others into *other*. Some small objects (e.g., cars) were also incorrectly assigned to the *vegetation* class. One reasonable explanation is that the structures of those objects are extremely small in the ALS scene so that they lose the main structural and morphological characteristics when represented by limited points.
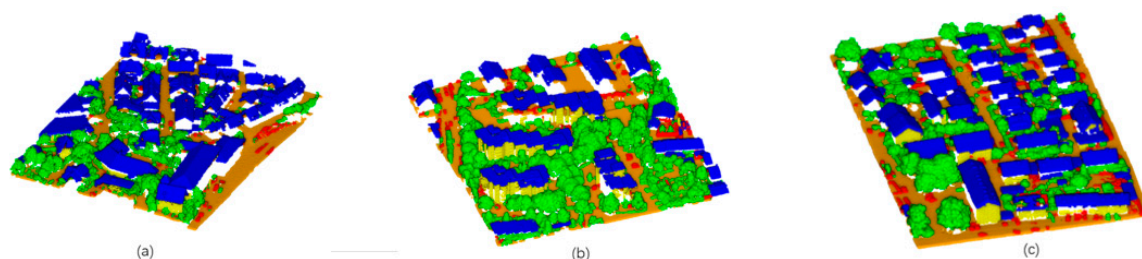

Figure 7. 3D View of the semantic labeling results for the three Vaihingen areas with five classes: *ground* (orange), *building roof* (blue), *vegetation* (green), *façade* (yellow), and *other* (red). (**a**) vs. Area 1, (**b**) vs. Area 2, and (**c**) vs. Area (3).

To effectively and objectively evaluate our methods, we introduced three indicators from the ISPRS suggestion for quantitative assessment of the qualities of the semantic labeling results [36]:

$$\text{Completeness} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{8}$$

$$\text{Correctness} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{9}$$

$$\text{Quality} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \tag{10}$$

where TP = true positives, FN = false negatives, and FP = false positives. The borders of the building roof were extracted from the semantic labeling results and then regularized. We exploited the regularized borders to implement a comparison with the benchmark. After two datasets (i.e., ours and the reference) were aligned in 2D space, we separately measured the areas of TP, FN, and FP. Figure 8 shows the evaluation of the semantic labeling results using the proposed method. Generally, there are only two obvious "missing" cases in the results (i.e., in Area 2 and Area 3, see more details in Figure 9) with areas larger than 50 m^2. However, such cases are due to the very poor quality of the point clouds, with respect to obtaining a regularized boundary, (e.g., the "missing" cases in Area 3, see Figure 9b) and the threshold setting (e.g., the "missing" case in Area 2 that is too close to the ground, <1 m, see Figure 9a).

Table 4 provides the performance of the quantitative assessment and comparisons with other classification methods (designated by acronyms representing the referenced method) in terms of the three indicators of completeness, correctness, and quality. From Table 4 we can see that in terms of completeness and correctness, the proposed method performs well and can obtain at least 90% of the

score. In correctness, the proposed method gives an excellent performance—-ranking number one in two of the three areas (see Table 4). Although in Area 3 VSK (98.7%) produces slightly better results than does our approach (98.3%), in terms of correctness, this performance is at the expense of lower completeness in VSK (86.3%). To attain a good score in correctness, we need to sacrifice a portion of completeness. The reason for this trade-off is that, in most cases, satisfactory correctness is attained at the cost of implementing an under-segmentation strategy. Thus, one of the important purposes of investigating classification methods is to balance these two factors, and the proposed method performs well in this respect. We can see that only our approach and that of HANC3 can keep those two scores at 90% or better in all three areas. In terms of completeness, HANC3 and our approach are both excellent in all areas, whereas in correctness, our approach performs better than HANC3, reducing the error rate by 25.5%, 25%, and 51.5% in Areas 1, 2, and 3, respectively. In quality, we highlight the top three group with bold italics. We can find that only our results remain in the top three group for all three areas in terms of quality. One main conclusion can be made from these experiments, that the proposed method is robust and the provides best balance between completeness and correctness, therefore performing well in a comprehensive evaluation.
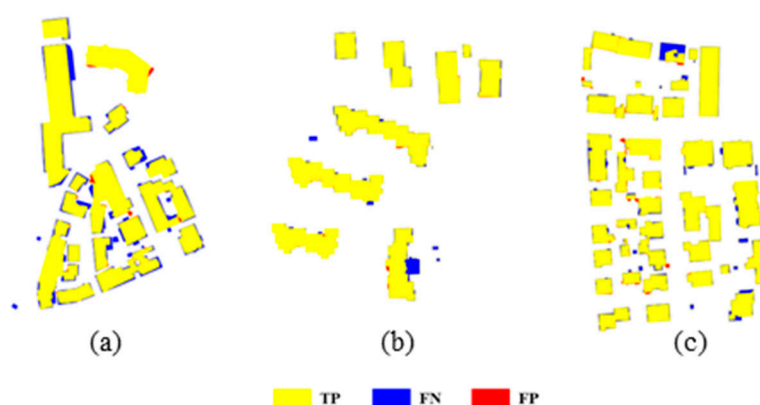

Figure 8. Semantic labeling results and visualized evaluation of three Vaihingen test sites. (**a**) vs. Area 1, (**b**) vs. Area 2, and (**c**) vs. Area 3.
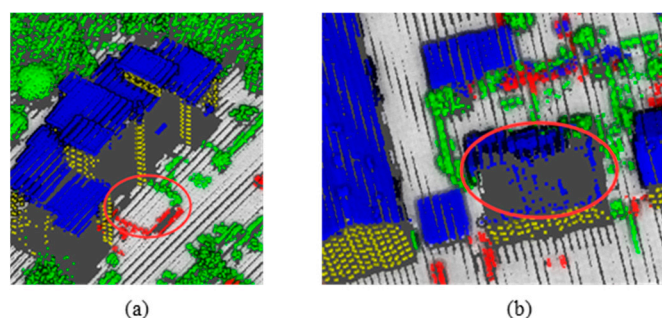

Figure 9. Two "missing" cases, where the areas are larger than 50 m^2 in Area 2 and Area 3 in the red circles of (**a**) and (**b**), respectively.

Table 4. Evaluations of the semantic labeling results in three Vaihingen sites using the proposed method and the quantitative comparison with other methods.

Method	AREA 1			AREA 2			AREA 3		
	Completeness %	Correctness %	Quality %	Completeness %	Correctness %	Quality %	Completeness %	Correctness %	Quality %
Ours	90.0	98.6	*88.9*	92.9	99.1	*92.1*	90.0	98.3	*88.6*
HKP [3]	92.0	97.4	*89.8*	93.0	98.4	*91.6*	89.2	97.7	87.4
HANC3 [5]	90.8	94.5	86.2	91.4	96.4	88.4	91.6	96.7	*88.8*
WHUY2 [37]	89.7	89.6	81.2	90.0	93.9	85.0	89.4	89.1	80.6
WHU_YD [38]	91.8	98.6	*90.6*	87.3	99.0	86.5	90.2	98.1	*88.6*
WHU_ZZ [39]	90.1	96.5	87.3	91.3	95.8	87.8	86.1	94.9	82.3
MON2 [40]	88.1	90.0	80.2	87.1	94.0	82.5	87.7	89.0	79.1
FED_2 [41]	85.4	86.6	75.4	88.8	84.5	76.4	89.9	84.7	77.3
TUM [42]	89.8	92.2	83.5	92.5	93.9	87.3	86.8	92.5	81.1
VSK [43]	85.7	98.1	84.3	85.4	98.4	84.2	86.3	98.7	85.3
MAR2 [44]	90.3	91.7	83.5	89.9	97.8	88.1	88.9	96.2	85.9
ITCR [6]	91.2	90.3	83.1	94.0	89.0	84.2	89.1	92.5	83.1
CAL2 [45]	87.7	97.2	85.5	90.7	96.7	88.0	89.2	97.7	87.4
DLR [46]	91.9	95.4	88.0	94.3	97.0	*91.6*	93.7	95.5	*89.7*

Note:Those methods' abbreviations are consistent with the benchmark tests in WG II/4 of ISPRS commissions II. http://www2.isprs.org/commissions/comm3/wg4/tests.html.

4.2. Hong Kong Dataset

The Hong Kong dataset covers an area larger than 3 km², with a length and width larger than 3 km and 1 km, respectively. Figure 10 shows the tiled image of this area. It is clear that the complexity of this scanning scene is far greater than that of the Vaihingen data in terms of the density of buildings and the architectural morphology of structures. Whole point clouds were already manually labeled with five relevant categories (i.e., *ground*, *building*, *low vegetation*, *high vegetation*, and *other*) by the Hong Kong government. Since the structures were so complex and irregular, a building was no longer separated into roof and façade parts. Moreover, the viaducts and footbridges in this dataset were also labeled as *building*. Here we should mention that although classifying the viaducts and footbridges as *building* will bias the classification in terms of the real environment, it does not bias the results of the experiments since we used the same categories in the training data. Furthermore, many old houses surrounded by vegetation cannot be well separated in the ALS point clouds cases even by a human operator. Thus, to find and separate vegetation is a big challenge in the semantic labeling of cases in such a dataset. The vegetation has been divided into two categories: *low vegetation* and *high vegetation*. In our experiments, we selected a specific part of the area to train the model. Figure 11 shows the whole area in the point clouds with the semantic labels: *ground* (orange), *building* (blue), *low vegetation* (light green), *high vegetation* (dark green), and *other* (red). The area in the red square is the part selected for training (including 2,352,724 labeled points), and the rest is used to test the proposed method (including 4,262,536 points).

Figure 10. Tiled image of the study area (red dotted square) in Hong Kong's Central District.
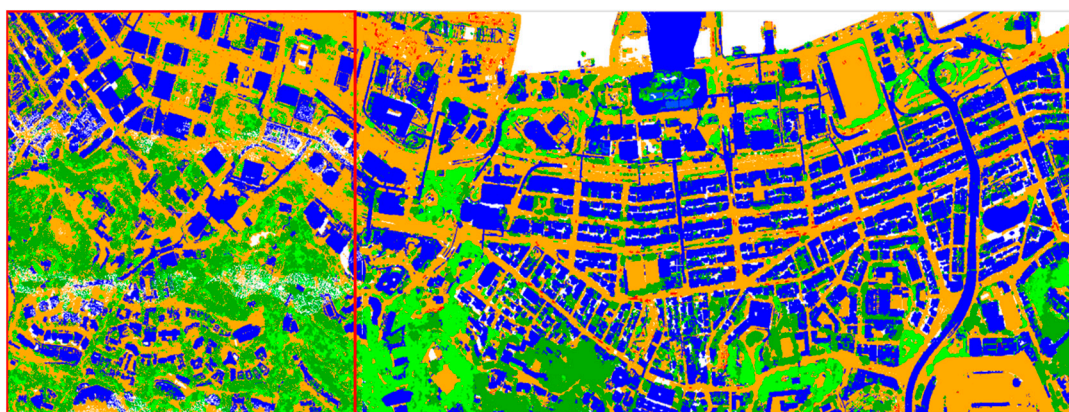

Figure 11. The training dataset and the semantic labeling results in 2D. The training dataset with manual labels is in the red square. The region outside the square comprises the semantic labeling results with labels as predicted by the proposed method.

In Figure 11, we display the semantic labeling results using the proposed approach in the test area. Figure 12 shows the test area in 3D, as part of the semantic labeling results. From Figures 11 and 12, we can first ascertain that the main buildings—-namely the areas larger than 50 m^2—-can be correctly detected. As previously mentioned, some large apparatus was assigned to *building* in the training dataset so that in the construction site we find a sprinkling of *building* objects (top left in Figure 11). The main vegetation areas are also found using the proposed method. In this case, it is too difficult to clearly identify all vegetation regions, even using a human operator, because some small vegetation areas appear as big as a bus within so few limited points.
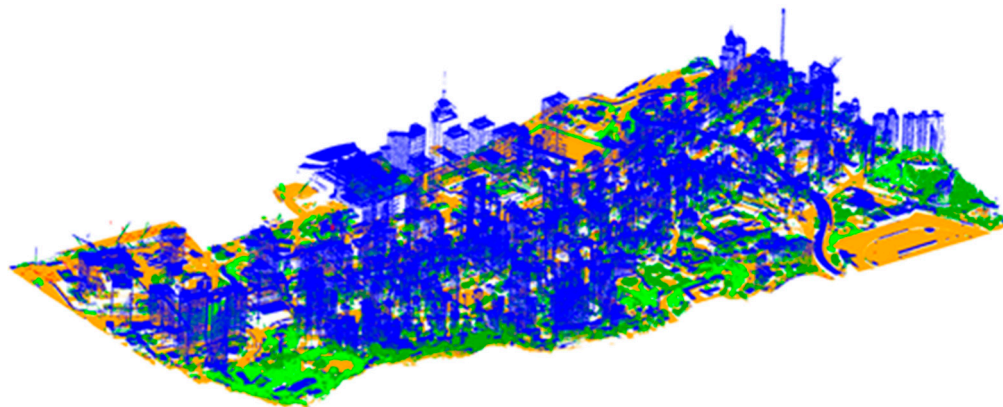

Figure 12. The semantic labeling results of the Hong Kong dataset using the proposed method in 3D.

As in our approach to the Vaihingen dataset, we first extracted the *building* class from the whole point clouds and then implemented an evaluation. Figure 13 shows the semantic labeling results evaluation. As mentioned earlier, most of the buildings can be safely detected by the proposed method. There is only one main building (larger than 50 m^2) that is missing from the results. That particular building area was filtered out as the *ground* class in the first step. Although there are many small parts missing, the score climbs to 89.2% in terms of completeness. The correctness of this test attains 95.6%. We also implemented a multi-primitive-based method without the proposed hierarchical optimal strategy and a multi-scaled point-based classification method on the Hong Kong dataset. The results are in Table 5. Clearly, the proposed method performs best out of all the tested methods. In addition, the performance of the proposed method (and, to a certain degree, the other two methods) is better than the results from the quantitative indicators in this case, because some correct results were identified as FP; this was caused by a wrongly, manually labeled dataset (see Figure 13; some footbridges in black squares should be assigned the *building* label based on the predefinition). Furthermore, there are many, very small isolated islands in the results (small red areas in the green squares in Figure 13) that belong to large double-decker buses but are assigned the *building* label. This problem can be addressed by adding a new class (i.e., car) in the training data for future investigation. The quality of the semantic labeling results climbs to 85.7%, confirming that the main layout of the buildings in the test area is correct.

Table 5. Evaluations of the semantic labeling results of the Hong Kong sites using the proposed method. MPB: multi-primitive-based; HS: hierarchical strategy.

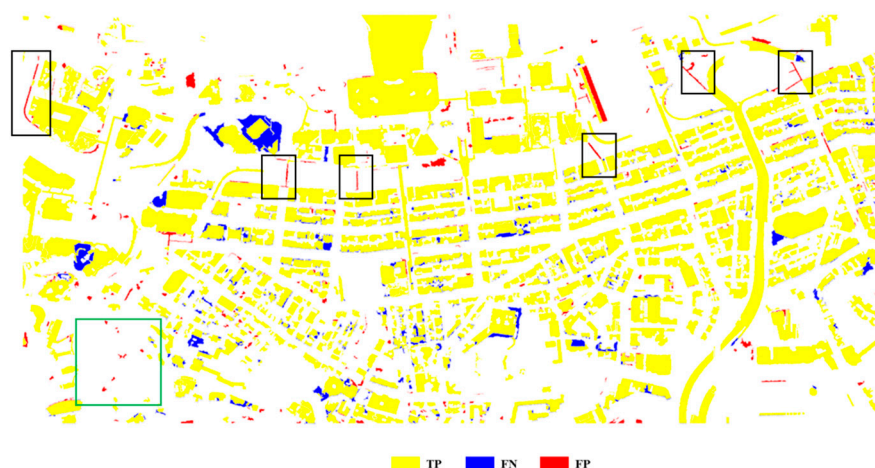| Method	Completeness	Correctness	Quality
MPB+HS (the proposed method)	89.2%	95.6%	85.7%
MPB	80.5%	88.4%	72.8%
Multi-scaled point-based method	75.9%	86.6%	67.9%

Figure 13. Semantic labeling results and visualized evaluation of the Hong Kong test sites.

5. Conclusions

This paper proposes a multi-primitive-based hierarchical optimal approach to labeling ALS point clouds. The input point clouds are appropriately segmented into and discriminative features are calculated for different primitives. From the existing relationships between different segments and primitives, we further categorize those relationships into two grades. Based on the initial semantic labeling results from the RFs, we carry out a hierarchical optimal strategy to smooth the semantic labeling.

Two challenging datasets are used to test and evaluate the properties of the proposed method. We find that our method strikes a reasonable balance between the two competing evaluation factors, completeness and correctness. The scores for correctness attained over 98% in all three cases of the ISPRS benchmark dataset and rank best in two out of the three benchmark datasets, with the remaining one having a score slightly lower than that of the best solution. More importantly, even with such a good performance, the results do not lose their completeness. Thus, the proposed method is one of the best in terms of the quality found in all three tests. The Hong Kong dataset also reflects the robustness of our method. The completeness and correctness attain approximately 89% and 96%, respectively, and quality achieves a high of 86%. From our comparisons, we find that the main proposed strategies, i.e., the multi-primitive-based and the hierarchical optimal methods perform exceptionally well. Our results could be used to detect small, detailed objects and be further exploited to support 3D city modeling or a city GIS system. Work to overcome some of the limitations of the proposed method could be conducted in the future, e.g., how to increase the considered number of classes.

Author Contributions: Investigation, X.G., B.W., L.Y., and H.H.; methodology, X.G.; software, X.G.; writing—original draft, X.G.; writing—review and editing, X.G. and B.W.

References

1. Chehata, N.; Guo, L.; Mallet, C. Airborne lidar feature selection for urban classification using random forests. International Archives of Photogrammetry. *Remote Sens. Spat. Inf. Sci.* **2009**, *38*(Part 3), W8.

2. Vosselman, G.; Coenen, M.; Rottensteiner, F. Contextual segment-based classification of airborne laser scanner data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *128*, 354–371. [CrossRef]

3. Zhu, Q.; Li, Y.; Hu, H.; Wu, B. Robust point cloud classification based on multi-level semantic relationships for urban scenes. *ISPRS J. Photogramm. Remote Sens.* **2017**, *129*, 86–102. [CrossRef]

4. Xu, S.; Vosselman, G.; Elberink, S.O. Multiple-entity based classification of airborne laser scanning data in urban areas. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 1–15. [CrossRef]

5. Niemeyer, J.; Rottensteiner, F.; Soergel, U. Contextual classification of lidar data and building object detection in urban areas. *ISPRS J. Photogramm. Remote Sens.* **2014**, *87*, 152–165. [CrossRef]

6. Gerke, M.; Xiao, J. Supervised and unsupervised MRF based 3D scene classification in multiple view airborne oblique images. ISPRS Annals of the Photogrammetry. *Remote Sens. Spat. Inf. Sci.* **2013**, *2*, 25–30.

7. Frey, B.J.; MacKay, D.J. A Revolution: Belief Propagation in Graphs with cycles. In Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems 10, Denver, CO, USA, 1997.

8. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *Ieee Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1222–1239. [CrossRef]

9. Niemeyer, J.; Rottensteiner, F.; Soergel, U. Conditional random fields for lidar point cloud classification in complex urban areas. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *3*, 263–268. [CrossRef]

10. Weinmann, M.; Schmidt, A.; Mallet, C.; Hinza, S.; Rottensteinerb, F.; Jutzia, B. Contextual classification of point cloud data by exploiting individual 3D neigbourhoods. ISPRS annals of the photogrammetry. *Remote Sens. Spat. Inf. Sci.* **2015**, *2*, 271.

11. Ni, H.; Lin, X.; Zhang, J. Classification of ALS point cloud with improved point cloud segmentation and random forests. *Remote Sens.* **2017**, *9*, 288. [CrossRef]

12. Darmawati, A.T. Utilization of Multiple Echo Information for Classification of Airborne Laser Scanning Data; Master's Thesis, International Institute for Geo-Information Science and Earth Observation (ITC), Enschede, The Netherlands, 2008.

13. Luo, H.; Wang, C.; Wen, C.; Chen, Z.; Zai, D.; Yu, Y.; Li, J. Semantic Labeling of Mobile LiDAR Point Clouds via Active Learning and Higher Order MRF. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3631–3644. [CrossRef]

14. Gevaert, C.M.; Persello, C.; Vosselman, G. Optimizing multiple kernel learning for the classification of UAV data. *Remote Sens.* **2016**, *8*, 1025. [CrossRef]

15. Niemeyer, J.; Rottensteiner, F.; Sörgel, U.; Heipke, C. Hierarchical higher order crf for the classification of airborne lidar point clouds in urban areas. The International Archives of Photogrammetry. *Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 655.

16. Mallet, C.; Bretar, F.; Roux, M.; Soergel, U.; Heipke, C. Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, S71–S84. [CrossRef]

17. Guo, B.; Huang, X.; Zhang, F.; Sohn, G. Classification of airborne laser scanning data using JointBoost. *ISPRS J. Photogramm. Remote Sens.* **2015**, *100*, 71–83. [CrossRef]

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

19. Landrieu, L.; Raguet, H.; Vallet, B.; Mallet, C.; Weinmann, M. A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 102–118. [CrossRef]

20. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. Pointnet: Deep learning on point sets for 3d classification and segmentation. *arXiv* **2016**, arXiv:1612.00593.

21. Landrieu, L.; Simonovsky, M. Large-scale point cloud semantic segmentation with superpoint graphs. *arXiv* **2017**, arXiv:1711.09869.

22. Hackel, T.; Wegner, J.D.; Savinov, N.; Ladicky, L.; Schindler, K.; Pollefeys, M. Large-Scale Supervised Learning For 3D Point Cloud Labeling: Semantic3d. Net. *Photogramm. Eng. Remote Sens.* **2018**, *84*, 297–308. [CrossRef]

23. Zhang, K.; Chen, S.; Whitman, D.; Shyu, M.L.; Yan, J.; Zhang, C. A progressive morphological filter for removing nonground measurements from airborne LIDAR data. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 872–882. [CrossRef]

24. Liu, K.; Boehm, J. A new framework for interactive segmentation of point clouds: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives. *Int. Soc. Photogramm. Remote Sens.* **2014**, *40*, 357–362. [CrossRef]

25. Schnabel, R.; Wahl, R.; Klein, R. Efficient RANSAC for Point-Cloud Shape Detection. *Computer Graphics Forumn 26.* **2007**, *2*, 214–226. [CrossRef]

26. Xu, B.; Jiang, W.; Shan, J.; Zhang, J.; Li, L. Investigation on the weighted ransac approaches for building roof plane segmentation from lidar point clouds. *Remote Sens.* **2016**, *8*, 5. [CrossRef]

27. Lin, Y.; Wang, C.; Cheng, J.; Chen, B.; Jia, F.; Chen, Z.; Li, J. Line segment extraction for large scale unorganized point cloud. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 172–183. [CrossRef]

28. Vosselman, G.; Klein, R. Visualisation and structuring of point clouds. In *Airborne and Terrestrial Laser Scanning*; Whittles Publishing: Dunbeath, UK, 2010; pp. 43–79.

29. Mei, J.; Zhang, L.; Wang, Y.; Zhu, Z.; Ding, H. Joint Margin, Cograph, and Label Constraints for Semisupervised Scene Parsing From Point Clouds. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3800–3813. [CrossRef]

30. Niemeyer, J.; Rottensteiner, F.; Sörgel, U.; Heipke, C. Contextual classification of point clouds using a two-stage CRF. The International Archives of Photogrammetry. *Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 141.

31. Wegner, J.D.; Montoya-Zegarra, J.A.; Schindler, K. A higher-order CRF model for road network extraction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1698–1705.

32. Delong, A.; Osokin, A.; Isack, H.N.; Boykov, Y. Fast approximate energy minimization with label cost. *Int. J. Comput. Vision* **2012**, *96*, 1–27. [CrossRef]

33. Chen, C.; Liaw, A.; Breiman, L. *Using Random Forest to Learn Imbalanced Data*; Technical Report 666; Statistics Department of University of California at Berkeley: Berkeley, CA, USA, 2004.

34. Rottensteiner, F.; Sohn, G.; Jung, J.; Gerke, M.; Baillard, C.; Benitez, S.; Breitkopf, U. The ISPRS benchmark on urban object classification and 3D building reconstruction. ISPRS Ann. Photogramm. *Remote Sens. Spat. Inf. Sci.* **2012**, *1*, 293–298.

35. Cramer, M.; Grenzdörffer, G.; Honkavaara, E. In-situ digital airborne camera validation and certification—The future standard? In Proceedings of the ISPRS 2010 Canadian Geomatics Conference and Symposium of Commission I, Calgargy, AB, Canada, 15–18 June 2010; Volume 38.

36. Rutzinger, M.; Rottensteiner, F.; Pfeifer, N. A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2009**, *2*, 11–20. [CrossRef]

37. Yang, B.; Xu, W.; Dong, Z. Automated extraction of building outlines from airborne laser scanning point clouds. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1399–1403. [CrossRef]

38. Yang, B.; Huang, R.; Li, J.; Tian, M.; Dai, W.; Zhong, R. Automated reconstruction of building LoDs from airborne LiDAR point clouds using an improved morphological scale space. *Remote Sens.* **2016**, *9*, 14. [CrossRef]

39. Zhao, Z.; Duan, Y.; Zhang, Y.; Cao, R. Extracting buildings from and regularizing boundaries in airborne lidar data using connected operators. *Int. J. Remote Sens.* **2016**, *37*, 889–912. [CrossRef]

40. Awrangjeb, M.; Lu, G.; Fraser, C. Automatic building extraction from LiDAR data covering complex urban scenes. The International Archives of Photogrammetry. *Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 25.

41. Gilani, S.; Awrangjeb, M.; Lu, G. Fusion of LiDAR data and multispectral imagery for effective building detection based on graph and connected component analysis. The International Archives of Photogrammetry. *Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 65.

42. Wei, Y.; Yao, W.; Wu, J.; Schmitt, M.; Stilla, U. Adaboost-based feature relevance assessment in fusing lidar and image data for classification of trees and vehicles in urban scenes. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *7*, 323–328. [CrossRef]

43. Dorninger, P.; Pfeifer, N. A comprehensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds. *Sensors* **2008**, *8*, 7323–7343. [CrossRef]

44. Mongus, D.; Lukač, N.; Žalik, B. Ground and building extraction from LiDAR data based on differential morphological profiles and locally fitted surfaces. *ISPRS J. Photogramm. Remote Sens.* **2014**, *93*, 145–156. [CrossRef]

45. Moussa, A. Segmentation and Classification of Multi-Sensor Data Using Artificial Intelligence Techniques. Ph.D. Thesis, University of Calgary, Calgary, AB, Canada, 2014.

46. Poznanska, A.M.; Bayer, S.; Bucher, T. Derivation of urban objects and their attributes for large-scale urban areas based on very high resolution UltraCam true orthophotos and nDSM: A case study. In Proceedings of the Earth Resources and Environmental Remote Sensing/GIS Applications IV, Berlin, Germany, 24 October 2013; Volume 8893, p. 889303.