CLASSIFICATION OF AERIAL POINT CLOUDS WITH DEEP LEARNING

E. Özdemir1,2, F. Remondino2
1 Space Center, Skolkovo Institute of Technology (SKOLTECH), Moscow, Russia
2 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy
Email: (cozdemir, remondino)@fbk.eu, web: http://3dom.fbk.eu

Commission II, WG II/4

KEY WORDS: point clouds, segmentation, classification, 3D building reconstruction

ABSTRACT:
Due to their usefulness in various implementations, such as energy evaluation, visibility analysis, emergency response, 3D cadastral, urban planning, change detection, navigation, etc., 3D city models have gained importance over the last decades. Point clouds are one of the primary data sources for the generation of realistic city models. Beside model-driven approaches, 3D building models can be directly produced from classified aerial point clouds. This paper presents an ongoing research for 3D building reconstruction based on the classification of aerial point clouds without given ancillary data (e.g. footprints, etc.). The work includes a deep learning approach based on specific geometric features extracted from the point cloud. The methodology was tested on the ISPRS 3D Semantic Labeling Contest (Vaihingen and Toronto point clouds) showing promising results, although partly affected by the low density and lack of points on the building facades for the available clouds.

Figure 1. The proposed classification pipeline applied to ALS data (~6pts/sqm) over a part of Toronto, Canada (ISPRS benchmark dataset - Niemeyer et al., 2014; Rottensteiner et al., 2014): (a) original data, (b) geometric feature extraction (local planarity shown), (c) DEM extraction, (d) classification with Deep Learning. The point cloud in (d) is colored as ground level objects - GLOs (blue), roofs (green) and vegetation (red).

1. INTRODUCTION

Over the last decades, 3D city models gained importance due to their wide range of applicability, in many use cases (i.e. visibility analysis, 3D cadastral, urban planning, etc.) with more than 100 applications (i.e. surveillance network planning, property registration, designing parks, etc.) (Biljecki et al., 2015). As 3D city models are being used for various purposes, it became a multi-disciplinary hot research field, as well. For this reason generation of 3D city models from photogrammetric or ALS point clouds has been studied by different researchers with different approaches (Haala and Kada, 2010; He et al., 2012; Lafarge and Mallet, 2012; Sampath and Shan, 2010; Toschi et al., 2017; Wagner et al., 2017). These works are often performed relying on external ancillary data, such as building footprints, DSM, DTM, etc (Fig. 2). However, it is not always possible to access such data, and even if they are reachable, they are not always up-to-date or correct, matching resolution and accuracy. These aforementioned reasons pushed us to develop a method for 3D building modeling based on a classified point cloud (Fig. 1), which enables to extract all needed information from one and only one input. Our aim is to use points geometric features in order to extract buildings’ roofs and facades, ground level objects (GLOs) and vegetation from aerial point clouds. This extracted information can be used afterwards for 3D building modeling without depending on any ancillary data.

This paper reports progresses to our previously presented work (Özdemir and Remondino, 2018). These advancements include deep learning implementation, removal of orthophoto processing, usage of only point cloud data, inclusion of more features in the classification workflow and improvement of classification with separation of building facades and roofs.

In the following sections, after a review of related works (Section 2), the developed method is reported in Section 3. Results are given in Section 4 before conclusions of the study in Section 5.

2. RELATED WORK

Many researches were performed on 3D building reconstruction using dense point clouds, as dense point clouds became more and more available in the last decades with the advances in LiDAR sensors and photogrammetry (Remondino et al., 2014). Most of these studies exploit available data (e.g. building footprints, DTM, DEM) in order to extract roofs, and then fitting geometric
primitives to these extracted roofs (Dorninger and Pfeifer, 2008; Holzmann et al., 2017; Li et al., 2019; Malihi et al., 2016; Vosselman and Dijkman, 2001; Xiong et al., 2014).

Figure 2. State-of-the-art 3D building reconstruction approaches.

Since our approach (Fig. 3) is based on extracting all needed information from a point cloud through classification, in the following sections we briefly share state-of-the-art methods related to the steps of our approach: point cloud segmentation, 3D geometric features and 3D deep learning.

Figure 3. The proposed approach to classify aerial photogrammetric point clouds using a deep learning method.

2.1 Point Cloud Segmentation and Classification

Point cloud segmentation has been an active research field for years, since it is a challenging task due data complexity. There have been different approaches developed during the years (Nguyen and Le, 2013). While some approaches focus on primitive fitting with sample consensus algorithm (Fischler and Bolles, 1981), some others rely on data fusion with combination of image and 3D data (Adam et al., 2018), in addition to the geometry based approaches, such as region growing (Rusu, 2010).

Similarly, classification of point clouds has always been a challenging task. In order to solve this complex problem, researches preferred to focus either on LiDAR or photogrammetric clouds. On the LiDAR side, Douillard et al. (2011) focused on a voxelization based approach for different kind of LiDAR scanners, Machet et al. (2017) developed a workflow to classify indoor LiDAR point clouds for Building Information Modeling (BIM) applications, Ramiya et al. (2017) proposed a classification method based on segmentation and histogram analysis. On the photogrammetric side, Becker et al. (2017) developed a machine learning based classification method - including both colour and geometry-based features. Dorninger and Nothegger (2007) developed a generic method that can handle unstructured point clouds, regardless of the data source.

2.2 Geometric Features for Point Clouds

The literature for point cloud classification show that geometric features of points are commonly used to classify point clouds. Many approaches focus on geometric features that can be implemented for both photogrammetric and LiDAR point clouds (Hackel et al., 2016a; Hackel et al., 2016b; Hackel et al., 2017a; Thomas et al., 2018; Weinmann et al., 2013; Weinmann et al., 2015b; Weinmann et al., 2015a). On the other hand, there are many approaches concentrating only on LiDAR point clouds (Charaniya et al., 2004; Dohan et al., 2015; Lalonde et al., 2006; Niemeyer et al., 2011b; Niemeyer et al., 2011a). Such methods frequently take advantage of eigenvalues and eigenvectors in order to extract geometric features and utilize machine learning classifiers in order to classify point clouds with respect to their geometric feature vectors.

2.3 Point Cloud Classification with Deep Learning

Deep learning has been used in various fields (such as natural language processing, speech recognition, computer vision, image processing, point cloud classification, etc.), due to its ability to solve complex problems (Deng, 2014; Goodfellow et al., 2016). For point cloud classification, a deep learning method has some advantages with respect to machine learning, as it does not need either handcrafting features to summarize your data nor a suitable classifier designed for your goal.

Different deep learning approaches for point cloud classification were presented in the literature: voxel-grid based classification (Hackel et al., 2017b; Wu et al., 2015), superpoint graph structure for semantic segmentation (Landrieu and Simonovsky, 2018), capturing local structures (Qi et al., 2017) and contextual features (Youselfhussein et al., 2018), etc.

3. DATA AND METHODOLOGY

Our aim is to classify aerial point clouds in order to extract buildings’ facades and roofs, vegetation and GLOs. The proposed method has two main parts: (i) data preparation with feature extraction and (ii) classification with deep learning (Fig. 3). While the classification part is straightforward (Section 3.3), the data preparation part includes (Section 3.2): extraction of geometric features, 3D region growing segmentation and DEM extraction.

In order to improve classification’s accuracy, we do not only rely on extracted geometric features (Section 3.2), but we exploit the neighbouring region of each point (Section 3.3).

3.1 Employed data

The ISPRS 3D Semantic Labeling Contest dataset (Niemeyer et al., 2014) is used to test and validate the developed method. The dataset includes point cloud data acquired with a Leica ALS50 airborne laser scanner over Vaihingen (Germany). In the dataset, points are labelled in 9 classes as follows: powerline, low vegetation, impervious surfaces, cars, fence/hedge, roof, facade, shrub and tree (Fig. 4). The point density of the dataset is ~5pts/sqm. The training set contains 753,876 points whereas the evaluation set contains 411,722 points.

3.2 Segmentation, Feature and DEM Extraction from Point Cloud

In our classification workflow, four custom features are used (Section 3.2.1-4): local planarity, vertical angle, elevation change...
and height above ground. Additionally, a new approach for DEM extraction from point cloud using region growing segmentation and geometric features is proposed.

3.2.1 3D Region Growing Segmentation: For our classification pipeline we use 3D region growing segmentation. This existing algorithm was implemented in Point Cloud Library (Rusu and Cousins, 2011), and we take advantage of it for our DEM extraction step. The algorithm works using the points’ curvature values. Initially, to prepare the seeds list, it sorts the points with respect to their curvature values. Afterwards, it starts growing the region with the point which has the minimum curvature value (Fig. 5). The points, which are qualified for the region, are removed from the list of seeds (Rusu, 2010).

3.2.2 Local Planarity: this feature is created by fitting a plane to neighbouring points and then calculating the average distance of those points the plane (Fig. 6). For plane fitting, we utilize the RANSAC (Fischler and Bolles, 1981) implementation available in the Point Cloud Library (Rusu and Cousins, 2011). As this geometric feature gives a metric about how planar the surrounding of each point is, we used it for separation of non-planar regions (i.e. vegetation) from planar regions (i.e. façades and roofs).

3.2.3 Vertical Angle: using 3D surface normal values of each point, the angle between the xy-plane and the normal vector is computed (Fig. 7). Using this geometric feature, surfaces can be distinguished with respect to their orientation. For example, facades form more vertical surfaces compared to roads and roofs, which are more horizontal surfaces.

3.2.4 Elevation Change: to calculate this feature for a given 3D point, minimum and maximum elevation values of neighbouring points are searched. The difference between these two values is assigned as elevation change value (Fig. 8).
3.2.5 Further Features: other geometric features are extracted from the point cloud using CloudCompare: anisotropy, surface variation, sphericity and verticality. All these features are extracted using eigen values, except for the verticality, which relies on eigenvectors. For a complete description of these features, please refer to (Hackel et al., 2016b).

3.2.6 Digital Elevation Model Extraction: in order to calculate a “height above ground” feature for each point in the point cloud (section 3.2.7), an approximate elevation of the ground level is needed. A new method (Fig. 9) is proposed to extract a hypothetical ground level with a DEM, utilizing region growing segmentation and geometric features (section 3.2.1).

As the aim is to extract a DEM in the final step, the region growing segmentation is adapted not to be too sensitive to small curvature (Fig. 10). This adjustment ensures two benefits: (i) an easier elimination of small details and (ii) final larger segments. Having separated the point cloud into segments, we initiate the suitable segment selection process with the largest segment, assuming it belongs to GLOs class.

While analysing the segments, we control their neighbouring points’ elevation differences and average elevation difference between segments (Fig. 11).

Following this step, the extracted point cloud is rasterized to form a DEM (Fig. 12), which is then used for the calculation of the “height above ground” feature (section 3.2.7).

3.2.7 Height Above Ground: starting from a DEM (Section 3.2.5), for each point in the point cloud, the elevation difference from the closest DEM point is calculated and assigned to the point (Fig. 13).

3.3 Training and Classification with Deep Learning

We utilized a simple deep learning approach based on a neural network with five layers including: sequence input layer, bidirectional long short-term memory layer with 200 hidden units, fully connected layer, softmax layer and classification layer (Fig. 14).

The reason to use a sequential deep learning approach is that we prefer to describe each point with its surrounding points, which can represent the geometry around the point in a better way compared to feature vectors. Therefore, for each point we sought its neighbouring points, and gave this as a sequence to the deep learning algorithm. For the training and classification purposes the following geometric features are used: height above ground, local planarity, anisotropy, surface variation, sphericity and verticality. In addition to the geometric features of the points in a sequence, we also included each point’s decentralizing the coordinates. In order to obtain these decentralized coordinates, we simply subtracted minimum x, y, z values within each sequence. Including both coordinates and geometric features in...
each sequence, our method can learn local geometric shapes and their variations for classification. Loss and accuracy of the network during training is shown in Figure 15.

![Figure 15. Loss and accuracy of the trained network for 4 classes.](image)

### 4. RESULTS

The proposed method is an improvement of (Özdemir and Remondino, 2018) which was developed assuming the availability of very dense point clouds and the need to generate 4 classes: roofs, façade, GLOs and vegetation. The Vaihingen dataset feature a low and varying point density, especially around the powerline, building façades and vegetation classes. This led to our feature extraction pipeline to eliminate many points (Table 1) as they do not have enough neighboring points for our algorithm within a certain search radius.

The Vaihingen dataset was processed in two runs: in the first run our deep learning algorithm (Section 3.3) was trained with the available 9 classes whereas in the second run it was trained with only 4 classes.

![Table 1. Number of points in each evaluation set for each class, before and after elimination.](image)

<table>
<thead>
<tr>
<th>Class</th>
<th>Reference Data</th>
<th>After Elimination</th>
<th>Lost Data</th>
<th>% Data Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powerline</td>
<td>600</td>
<td>98</td>
<td>502</td>
<td>84%</td>
</tr>
<tr>
<td>Low veget.</td>
<td>98690</td>
<td>93467</td>
<td>5223</td>
<td>5%</td>
</tr>
<tr>
<td>Imp. surf.</td>
<td>101986</td>
<td>97853</td>
<td>4133</td>
<td>4%</td>
</tr>
<tr>
<td>Car</td>
<td>3708</td>
<td>3235</td>
<td>473</td>
<td>13%</td>
</tr>
<tr>
<td>Fence</td>
<td>7422</td>
<td>7087</td>
<td>335</td>
<td>5%</td>
</tr>
<tr>
<td>Roof</td>
<td>109048</td>
<td>103897</td>
<td>5151</td>
<td>5%</td>
</tr>
<tr>
<td>Façade</td>
<td>11224</td>
<td>7533</td>
<td>3691</td>
<td>33%</td>
</tr>
<tr>
<td>Shrub</td>
<td>24816</td>
<td>23230</td>
<td>1588</td>
<td>6%</td>
</tr>
<tr>
<td>Tree</td>
<td>54226</td>
<td>47486</td>
<td>6740</td>
<td>12%</td>
</tr>
<tr>
<td>Total</td>
<td>411722</td>
<td>383886</td>
<td>27836</td>
<td>7%</td>
</tr>
</tbody>
</table>

For the 9-class classification of the evaluation data, results are shown in Figure 17, while confusion matrix and per-class accuracy are given in Table 2. For the second run, we merged or removed some classes in the training set in order to get 4 classes. The removed classes are powerline, car and fence. We merged tree and shrub classes to form vegetation class, and similarly, low vegetation and impervious surfaces to form the GLOs class. For the 4-class classification, results are shown in Figure 18 while confusion matrix and achieved accuracy results are given in Table 3. As previously mentioned, our objective is to classify the point cloud into 4 classes. Therefore, our geometric features are designed to emphasize the difference between these 4 classes but not all available 9 classes in the dataset. As we did not select our features to highlight powerlines, cars and fences, accuracies in 9-class classification are lower compared to 4-class classification.

Similarly, looking for the available results on the benchmark’s website, while our 9-class accuracies are lower than the vast majority of them, our 4-class accuracies are similar to average results.

In addition to the tests on the Vaihingen dataset, the replicability of the used training was evaluated, in order to see how it could classify other kind of data. This is indeed an important issue in the Geomatics and Geoinformatics field as many researchers are working on the classification problem but replicability and scalability are still very open issues. In order to solve these issues, a sensor-independent dataset, at least for roofs, can be acquired and tested with the proposed RootN3D (Wichmann et al., 2018).

The network trained with 4-classes (Table 3) was used to classify a small portion of ISPRS benchmark dataset of Dortmund City Center (Nex et al., 2015). This is a denser point cloud data with ~50pts/sqkm density, acquired using oblique photogrammetry technique, instead of laser scanner.

In the classification results (Fig. 16) it can be seen that facades could not be classified, whereas the GLOs have a proper look and roofs are arguably-correct classified. We assume this situation is mainly caused by the difference in point density between the two datasets, and the lack of points on vertical structures during the training procedure.

![Figure 16. A portion of the Dortmund benchmark dataset, original cloud part (top), and classified (bottom) with 4-class classifier: roofs (green), vegetation (red) and GLOs (blue).](image)

### 5. CONCLUSIONS

This paper reported an approach for point cloud classification for a successive 3D building reconstruction procedure. The developed automated procedure includes deep learning processing and the extraction of buildings’ roofs and facades, beside vegetation and GLOs. The procedure relies on geometric features, which are calculated for each 3D point with their neighboring points within a certain radius. The density of the available cloud plays a key role in the classification procedure: if points are sparse, geometric features cannot be properly calculated and such points need to be eliminated (Table 1). Moreover, an imbalanced loss of façade points in the training data caused misclassifications. Nevertheless, newly generated aerial dense point clouds are more dense, with oblique photogrammetric point clouds featuring points also on building façades (Nex et al., 2015). Therefore, extracting semantic classes from such point clouds will facilitate to derive 3D building geometries based on extracted entities.
**Table 2. Confusion Matrix and per-class accuracy for 9-class classification.**

<table>
<thead>
<tr>
<th>Class Name</th>
<th>Powerline</th>
<th>Low veget.</th>
<th>Imp. surf.</th>
<th>Car</th>
<th>Fence</th>
<th>Roof</th>
<th>Facade</th>
<th>Shrub</th>
<th>Tree</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
<th>Balanced Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powerline</td>
<td>36</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>3</td>
<td>31</td>
<td>36.7%</td>
<td>1.4%</td>
<td>2.70%</td>
<td>19.1%</td>
</tr>
<tr>
<td>Low veget.</td>
<td>0</td>
<td>22690</td>
<td>43360</td>
<td>3</td>
<td>826</td>
<td>10899</td>
<td>118</td>
<td>15427</td>
<td>144</td>
<td>24.3%</td>
<td>70.2%</td>
<td>36.07%</td>
<td>47.2%</td>
</tr>
<tr>
<td>Imp. surf.</td>
<td>0</td>
<td>5951</td>
<td>88366</td>
<td>0</td>
<td>196</td>
<td>2358</td>
<td>7</td>
<td>952</td>
<td>3</td>
<td>90.3%</td>
<td>65.5%</td>
<td>75.95%</td>
<td>77.9%</td>
</tr>
<tr>
<td>Car</td>
<td>0</td>
<td>200</td>
<td>9</td>
<td>3</td>
<td>548</td>
<td>433</td>
<td>2</td>
<td>2040</td>
<td>0</td>
<td>0.1%</td>
<td>42.9%</td>
<td>0.19%</td>
<td>21.5%</td>
</tr>
<tr>
<td>Fence</td>
<td>0</td>
<td>364</td>
<td>51</td>
<td>0</td>
<td>680</td>
<td>770</td>
<td>74</td>
<td>4271</td>
<td>877</td>
<td>9.6%</td>
<td>20.1%</td>
<td>12.98%</td>
<td>14.8%</td>
</tr>
<tr>
<td>Roof</td>
<td>1535</td>
<td>560</td>
<td>2479</td>
<td>0</td>
<td>339</td>
<td>93387</td>
<td>73</td>
<td>1601</td>
<td>3923</td>
<td>89.9%</td>
<td>80.8%</td>
<td>85.10%</td>
<td>85.3%</td>
</tr>
<tr>
<td>Facade</td>
<td>102</td>
<td>296</td>
<td>98</td>
<td>0</td>
<td>28</td>
<td>1271</td>
<td>2344</td>
<td>1650</td>
<td>1744</td>
<td>31.1%</td>
<td>76.6%</td>
<td>44.25%</td>
<td>53.8%</td>
</tr>
<tr>
<td>Shrub</td>
<td>0</td>
<td>1689</td>
<td>473</td>
<td>1</td>
<td>456</td>
<td>1775</td>
<td>125</td>
<td>15302</td>
<td>3409</td>
<td>65.9%</td>
<td>30.9%</td>
<td>42.04%</td>
<td>48.4%</td>
</tr>
<tr>
<td>Tree</td>
<td>892</td>
<td>587</td>
<td>27</td>
<td>0</td>
<td>315</td>
<td>4664</td>
<td>318</td>
<td>8313</td>
<td>32370</td>
<td>68.2%</td>
<td>76.2%</td>
<td>71.94%</td>
<td>72.2%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46.2%</td>
<td>51.6%</td>
<td>41.2%</td>
<td>48.9%</td>
</tr>
</tbody>
</table>

**Table 4. Confusion Matrix and per-class accuracy for 4-class classification. Others* include points from powerline, cars and fence, which are classified as GLOs, roof, facade or vegetation.**

<table>
<thead>
<tr>
<th>Class Name</th>
<th>GLO</th>
<th>Roof</th>
<th>Facade</th>
<th>Vegetation</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
<th>Balanced Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLO</td>
<td>169454</td>
<td>8380</td>
<td>561</td>
<td>12925</td>
<td>88.6%</td>
<td>94.6%</td>
<td>91.5%</td>
<td>91.6%</td>
</tr>
<tr>
<td>Roof</td>
<td>3404</td>
<td>92100</td>
<td>36</td>
<td>8357</td>
<td>88.6%</td>
<td>84.9%</td>
<td>86.7%</td>
<td>86.8%</td>
</tr>
<tr>
<td>Facade</td>
<td>608</td>
<td>1185</td>
<td>1738</td>
<td>4002</td>
<td>53.1%</td>
<td>65.1%</td>
<td>58.5%</td>
<td>59.1%</td>
</tr>
<tr>
<td>Vegetation</td>
<td>4418</td>
<td>5537</td>
<td>283</td>
<td>60478</td>
<td>85.5%</td>
<td>64.6%</td>
<td>73.6%</td>
<td>75.1%</td>
</tr>
<tr>
<td>Others*</td>
<td>1210</td>
<td>1289</td>
<td>52</td>
<td>7869</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>79.0%</td>
<td>77.3%</td>
<td>77.6%</td>
<td>78.1%</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS
The authors would like to acknowledge the provision of the Downtown Toronto data set by Optech Inc., First Base Solutions Inc., GeoICT Lab at York University, and ISPRS WG II/4. The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) (Cramer, 2010): http://www.ifp.uni-stuttgart.de/dgpf/DE/DEP-Allg.html.

REFERENCES


