DublinCity: Annotated LiDAR Point Cloud and its Applications

S. M. Iman Zolanvari[1]
https://zolanvari.com
Susana Ruano[1]
ruanosas@scss.tcd.ie
Aakanksha Rana[1]
ranaa@scss.tcd.ie
Alan Cummins[1]
alan.cummins@tcd.ie
Rogerio Eduardo da Silva[2]
www.rogerioesilva.net
Morteza Rahbar[34]
rahbar@arch.ethz.ch
Aljosa Smolic[1]
smolica@scss.tcd.ie

[1] V-SENSE
School of Computer Science and Statistics
Trinity College Dublin, Ireland
[2] University of Houston-Victoria, Victoria, Texas, US
[3] CAAD
Department of Architecture
ETH, Zurich, Switzerland
[4] Department of Architecture
Tarbiat Modares University
Tehran, Iran

Abstract

Scene understanding of full-scale 3D models of an urban area remains a challenging task. While advanced computer vision techniques offer cost-effective approaches to analyse 3D urban elements, a precise and densely labelled dataset is quintessential. The paper presents the first-ever labelled dataset for a highly dense Aerial Laser Scanning (ALS) point cloud at city-scale. This work introduces a novel benchmark dataset that includes a manually annotated point cloud for over 260 million laser scanning points into 100'000 (approx.) assets from Dublin LiDAR point cloud [12] in 2015. Objects are labelled into 13 classes using hierarchical levels of detail from large (*i.e.* building, vegetation and ground) to refined (*i.e.* window, door and tree) elements. To validate the performance of our dataset, two different applications are showcased. Firstly, the labelled point cloud is employed for training Convolutional Neural Networks (CNNs) to classify urban elements. The dataset is tested on the well-known state-of-the-art CNNs (*i.e.* PointNet, PointNet++ and So-Net). Secondly, the complete ALS dataset is applied as detailed ground truth for city-scale image-based 3D reconstruction.

1 Introduction

In computer vision, automated identification of three-dimensional (3D) assets in an unstructured large dataset is essential for scene understanding. In order to collect such a big dataset at city-scale, laser scanning technology, also known as Light Detection and Ranging (LiDAR), offers an efficient means of capture. Aerial Laser Scanning (ALS), a LiDAR data

acquisition approach, is generally used for obtaining data for a large area (*e.g.* an entire urban region). The data consists of unprocessed waveforms and a discrete point cloud. In addition, image data can be collected alongside this LiDAR point cloud data if an additional camera is used.

To develop algorithms which are able to automatically identify 3D shapes and objects in a large LiDAR dataset, it is crucial to have a well-annotated ground-truth data available. However, the generation of such labels is cumbersome and expensive. It is essential to have access to a full 3D, dense, and non-synthetic labelled point cloud at city-scale that includes a variety of urban elements (*e.g.* various types of roofs, buildings' facade, windows, trees and sidewalks). While there have been several attempts to generate such a labelled dataset by using photogrammetric or morphological methods, review of the well-known available datasets shows that none of these can completely satisfy all requirements [17]. Therefore, this paper presents a novel manually annotated point cloud from ALS data of Dublin city centre that was obtained by Laefer *et al.* [12]. This dataset is one of the densest urban aerial LiDAR datasets that have ever been collected with an average point density of 348.43 points/m^2 [8].

The main contribution of this paper is the manual annotation of over 260 million laser scanning points into 100'000 (approx.) assets into 13 classes (*e.g.* building, tree, facade, windows and streets) with the hierarchical levels. The proposed labelled dataset is the first of its kind regarding its accuracy, density and diverse classes, particularly with the city scale coverage area. To the best knowledge of the authors, no publicly available LiDAR dataset is available with the unique features of the DublinCity dataset. The hierarchical labels offer excellent potential for classification and semantic segmentation of urban elements from refined (*e.g.* windows and doors) to coarse level (*e.g.* buildings and ground). Herein, two different applications are introduced to validate the important usage of the labelled point cloud.

The first application is an automated technique for classification of 3D objects by using three state-of-the-art CNN based models. Machine Learning (ML) techniques offer relatively efficient and high accuracy means to process massive datasets [33]. ML techniques highly rely on input training datasets. Most datasets are not generated for detailed 3D urban elements on a city-scale. Therefore, there is limited research in that direction. Herein, three highly cited CNNs are trained and evaluated by employing the introduced manually labelled point cloud. Secondly, in the original dataset, aerial images were also captured during the helicopter fly-over. By using this image data, an image-based 3D reconstruction of the city is generated and the result is compared to the LiDAR data across multiple factors. The next section reviews previous related datasets and the aforementioned relevant applications in more detail.

2 Related Work

Several state-of-the-art annotated point cloud datasets have provided significant opportunities to develop and improve 3D analysis algorithms for various related research fields (*e.g.* computer vision, remote sensing, autonomous navigation and urban planning). However, due to varying limitations, such datasets are not ideal to achieve the necessary accuracy. Hence, the availability of high quality and large size point cloud datasets is of utmost importance.

In the ShapeNet dataset [38], point clouds are not obtained from scanning of real objects, rather they are generated from 3D synthetic CAD models. In TerraMobilita/iQmulus [35]

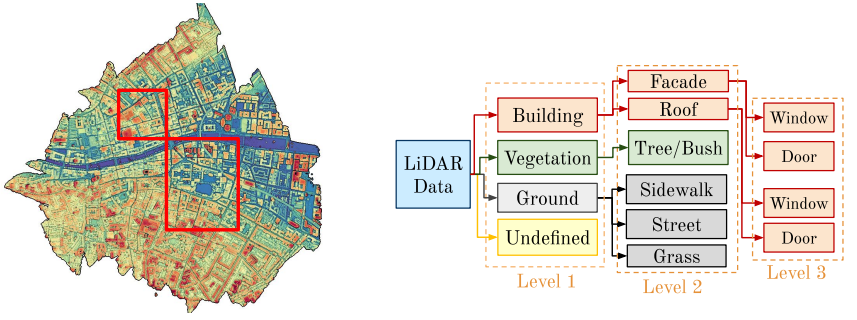
and more recently in the Paris-Lille [25] project, the datasets are obtained by Mobile Laser Scanning (MLS). However, MLS LiDAR datasets are not fully 3D, as MLS datasets can only scan the ground and buildings' facades without any information from the roof or the other sides of buildings. While the ScanNet [6] dataset consists of real 3D objects, it only includes indoor objects (*e.g.* chairs, desks and beds) without any elements from outside (*e.g.* buildings' facade or roofs). In contrast, the Semantic3D [9] dataset generated outdoor point cloud from several registered Terrestrial Laser Scanning (TLS). However, it only covers a small portion of a city with a limited number of elements. Similarly, RoofN3D [37] only includes a specific type of element (*i.e.* roofs) without even coverage of distinct urban roofs (*e.g.* flat, pyramid, shed and M-shaped). More recently, the TorontoCity [36] dataset was introduced which uses high-precision maps from multiple sources to create the ground truth labels. However, the dataset still has no manually labelled objects. In addition to these datasets, AHN 1,2 and 3 datasets [1] also provided large scale LiDAR data for a vast area of Netherlands. While the AHN datasets covered a large area, the average density of the point cloud is only around 8 to 60 points/m^2 which is not sufficient for generation of a detailed 3D model.

3D Point Cloud Object Classification. Classification using the 3D point cloud data has gained considerable attention across several research domains *e.g.* autonomous navigation [21, 24], virtual and augmented reality creation [10, 23] and urban[13, 34]-forest monitoring [19] tasks. Amongst the state-of-the-art classification techniques, CNN based models offer a reliable and cost-effective solution to process 3D point cloud datasets that are massive and unstructured in nature. The earliest CNN model was introduced by [16], where voxel grids are used as inputs. With rapid advancement in deep learning models, recently, several methods [12, 20, 22] have been proposed in the literature that utilise the point cloud data to train the CNN rather than the voxel grids [16, 40] or collections of images.

PointNet [20, 22] is one of the first successful attempts to describe the point cloud object using a distinctive global feature representation. As an extension to the former, PointNet++ [22] is designed to further address the 'fine' local details in objects by building a pyramid-like feature aggregation model. To achieve better performance, the recently proposed SO-Net model in [14] formulates the single feature vector representation using the hierarchical feature extraction on individual points.

Image-based 3D Reconstruction. Recently, driven by the collection of large imagery datasets, the classic computer vision challenge of image-based 3D reconstruction has been revisited to focus on large-scale problems [5, 26]. As a result, the list of open-source methods for Structure-from-Motion (SfM) [18, 26, 31, 33] and Multi-view Stereo (MVS) [4, 7, 28] has increased. However, one of the main issues working with these algorithms is the lack of ground truth for measuring the accuracy of the reconstruction, which is usually acquired with active techniques (*e.g.* LiDAR) and, therefore, it is rarely available for large-scale areas [11, 15].

There are several benchmarks available for 3D reconstruction. One of the first is the Middlebury benchmark [30] which has been recently extended [3]. In [3] database, the number of images and objects has increased. However, they provide less than 100 images for each reconstructed model and focus on small objects in a confined indoor space. The EPFL benchmark [32] and the ETH3D benchmark [29] are also presented to fill the gap between image-based reconstruction techniques and the LiDAR outdoors. While the benchmark dataset is acquired with terrestrial LiDAR, the number of images and models are limited and they only



(a) Overview of data and the selected area where labelled areas highlighted in red

(b) Hierarchy of classes (all 13 classes in 3 levels)

Figure 1: Overview of the database

focus on a few monuments. A dataset that overcomes the limitation of the terrestrial LiDAR is the Toronto/Vaihingen ISPRS used in [39], but it still covers a small area of the city compared to the aerial LiDAR dataset available for the city of Dublin which is presented in this work. In most recent benchmarks [11, 15, 29] COLMAP [26, 28] is reported as the best approach (SfM + MVS) to generate the reconstructions in the majority of the scenarios. Therefore, in this work we use it to generate the image-based reconstructions.

3 Database Specification

The initial dataset [12] includes a major area of Dublin city centre (*i.e.* around 5.6 km^2 including partially covered areas) was scanned via an ALS device which was carried out by helicopter in 2015. However, the actual focused area was around 2 km^2 which contains the most densest LiDAR point cloud and imagery dataset. The flight altitude was mostly around 300m and the total journey was performed in 41 flight path strips.

LiDAR data. The LiDAR point cloud used in this paper is derived from a registered point cloud of all these strips. Since the whole stacked point cloud includes more than 1.4 billion points, they are split into smaller tiles to be loaded and processed efficiently. The Dublin City airborne dataset is one of the world's densest urban ALS dataset ever collected. The final registered LiDAR point cloud offers an average density of 250 to 348 points/m^2 in various tiles. Figure 1a shows an overview of the whole LiDAR point cloud that is colourised with regard to the elevation of the points. The selected area for labelling is highlighted inside the red boxes.

In this paper, around 260 million points (out of 1.4 billion) were labelled. The selected area is within the most densely sampled part of the point cloud with full coverage by aerial images. This area (*i.e.* inside the red boxes) includes diverse types of historic and modern urban elements. Types of buildings include offices, shops, libraries, and residential houses. Those buildings are in the form of detached, semi-detached and terraced houses and belong to different eras (*e.g.* from 17th century rubrics building to the 21th century George's Quay complex as a modern structure).

In the initial data acquisition, generating LiDAR point cloud was the primary output of the aerial scanning. However, the helicopter also collected imagery data during the flight [12]. Two different sets of images are detailed:

Top view imagery. The dataset includes 4471 images taken from a helicopter and are

named in the cited repository as Geo-referenced RGB. The resolution in pixels of the images is 9000×6732 and are stored in TIFF format, with a ground sampling distance of 3.4 cm. The total size of the images is around 813 GB and they are presented in the different flight paths. The geographic information is given as a GPS tag in the EXIF metadata and the camera used for the capture is Leica RCD30.

Oblique view imagery. The oblique imagery contains 4033 JPEG images, which are taken with two different NIKON D800E cameras. They are referred to as Oblique photos, their resolution is 7360×4912 and its size is 18.5 GB. As per the geo-registered RGB images, they are presented in accordance with the flight path but include an extra subdivision which corresponds to each camera.

Manual Labelling Process. Herein, a manually annotated airborne LiDAR dataset of Dublin is presented. A subset of 260 million points, from the 1.4 billion laser point cloud obtained, have been manually labelled. The labels represent individual classes and they are included in three hierarchical levels (Figure 1b):

i. **Level 1:** This level produces a coarse labelling that includes four classes: (a) Building; (b) Ground; (c) Vegetation; and (d) Undefined. Buildings are all shapes of habitable urban structures (*e.g.* homes, offices, schools and libraries). Ground mostly contains points that are at the terrain elevation. The Vegetation class includes all types of separable plants. Finally, Undefined points are those of less interest to include as urban elements (*e.g.* bins, decorative sculptures, cars, benches, poles, post boxes and non-static object). Approximately 10% of the total points are labelled as undefined and they are mostly points of river, railways and construction sites.

ii. **Level 2:** In this level, the first three categories of Level 1 are divided into a series of refined classes. Buildings are labelled into roof and facade. Vegetation is divided into separate plants (*e.g.* trees and bushes). Finally, Ground points are split into street, sidewalk and grass.

iii. **Level 3:** Includes any types of doors and windows on roofs (*e.g.* dormers and skylights) and facades.

In order to label the LiDAR data, it is divided into smaller sub-tiles (*i.e.* each includes around 19 million laser scanning points). The process starts with importing data into the CloudCompare 2.10.1 [2] software. Then, points were coarsely manually segmented with segmentation and slicing tools in three categories (*i.e.* building, vegetation and ground) and labelled accordingly. Then, the process continues to the third level which has the finest details (*i.e.* windows and doors). Thereby, this pipeline produces a unique label for each point. The process is performed in over 2500 hours with an appropriate tutorial, supervision, and carefully crossed-checked multiple times to minimise the error. The naming order and the detail of labelling are demonstrated in more detail within the supplementary material. Figure 2 shows the visual representation of the classes for one sub-tile which consists of around 19 million laser scanning points.

4 Results and Applications

As shown in Figure 3a, the majority of captured points belong to the ground classes. This is expected because point cloud density for the horizontal planes reflects more points towards the aerial scanner in comparison to the vertical surfaces (*e.g.* facades). Also, windows have

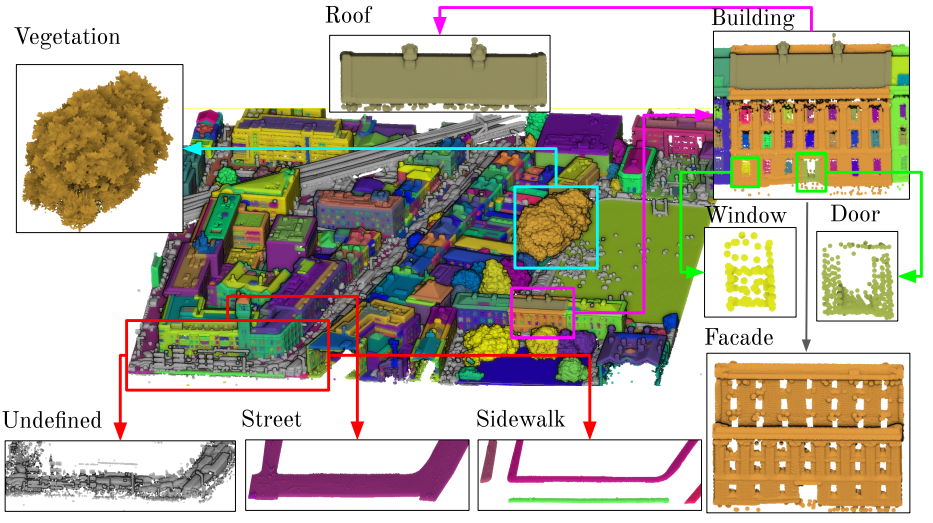
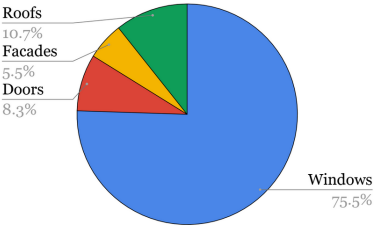


Figure 2: A Sample labelled sub-tile that visualises all classes

Classes		#Points (x10^3)	
Building	window	604	115'870
	facade	55'756	
	roof	59'134	
	door	376	
Vegetation	tree	19'071	19'071
Ground	sidewalk	22'257	102'758
	street	64'593	
	grass	15'908	

(a) Total number of labelled points for each class



(b) Percentage of assets only in building class

Figure 3: Overview of the labelled results

fewer points as laser beam usually penetrates the glass and there are only a few points on it. Similarly, the number of points in the door class is smaller as each building normally has one door for its entrance and one door for access to the roof. While this table shows the total number of points, Figure 3b shows the frequency percentage of each class that building category contains. For example, around 75% of the objects in the building category are windows and around 8.3% are doors as the number of windows in each building is much higher than other classes.

The manually annotated point cloud dataset is available at https://v-sense.scss.tcd.ie/DublinCity/. An additional video is provided in the attached supplementary material for visual demonstration of the dataset. In the next sections, two applications are showcased by employing the annotated point cloud dataset.

4.1 3D Point Cloud Object Classification

The performance analysis of the dataset for object classification problem is showcased by using the state-of-the-art models namely, PointNet [20], PointNet++ [22] and SO-Net [14].

These three models directly work on unstructured point cloud datasets. They learn the global point cloud features that have been shown to classify forty man-made objects of the ModelNet40 [33] shape classification benchmark.

The 3D point cloud dataset is comprised of a variety of outdoor areas (*i.e.* university campus and city centre) with structures of facades, roads, door, windows and trees as shown in Figure 2. In order to study the classification accuracy on the three CNN-based models, a dataset of 3982 objects of 5 classes (*i.e.* doors, windows, facades, roofs and trees) is gathered.

To evaluate the three models, the dataset is split into a ratio of 80 : 20 for training and testing respectively. While training, for each sample, points on mesh faces are uniformly sampled according to the face area and normalised into a unit sphere (*i.e.* -1 to +1). Additionally, data augmentation techniques are applied on-the-fly by randomly rotating the object along the up-axis and jittering the position of each point by Gaussian noise with zero mean and 0.02 standard deviation. Each model is trained for 100 epochs.

In Table 1, the performance of the three trained models in a different point cloud input setting using the Overall and Average class accuracy (as used in [20, 22]) is shown. It is observed that with an increase in the number of points per objects, the performance of the three models increases. Amongst all the three networks, the So-Net architecture performs the best. This is in consistence with the results in [14]. However, there is still a huge potential in the improvement of the performance scores. This is primarily because dataset is challenging in terms of structural similarity of outdoor objects in the point cloud space namely, facades, door and windows.

	PointNet [20]		**PointNet++** [22]		**So-Nets** [14]	
#Points	Avg. Class	Overall	Avg. Class	Overall	Avg. Class	Overall
512	24.17	35.17	39.47	45.56	41.89	48.74
1024	38.84	50.13	44.65	62.91	45.73	63.54
2048	46.77	59.68	49.23	63.42	49.34	64.55
4096	48.77	60.68	51.23	64.42	50.34	65.55

Table 1: Overall and Avg. class classification scores using the state-of-the-art models on the dataset.

4.2 Image-based 3D Reconstruction

In this section, the whole extension of the LiDAR point cloud described in Section 3 is exploited beyond the annotated data. To do that, an evaluation of the image-based 3D reconstruction is presented using two different types of aerial imagery data that are collected alongside the LiDAR data [12]. The first set is composed of images with a top view of the city (by a downward-looking camera) and the second set consists of oblique views (by a $\approx 45^\circ$ tilted camera). More details of the images are given in Section 3.

In order to carry out the experiment, the complete reconstruction pipeline is evaluated as per [11]. This is because the ground truth for the camera positions to specifically evaluate the camera poses from an SfM algorithm are not available, only the GPS position is known. The open-source software selected for the reconstruction is COLMAP (SfM [26] and MVS [28]), since it is reported as the most successful in different scenarios in the latest comparisons carried out (see Section 2). Furthermore, it has advantages over other methods [4, 18] because

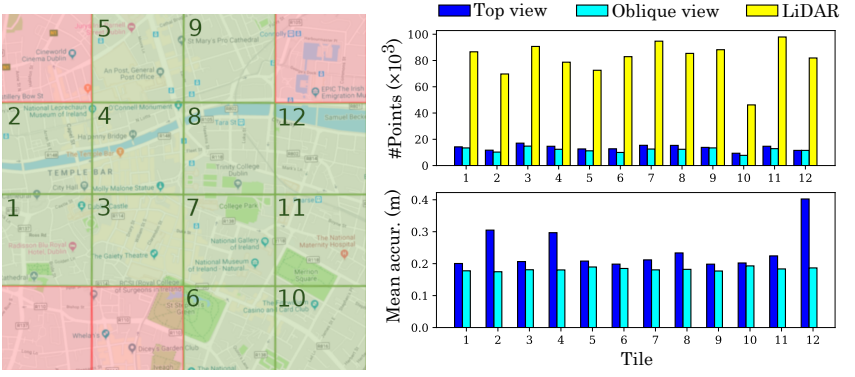


Figure 4: On the left, the area of the city covered by the tiles in the comparison (in green). On the right, a comparison of the number of points and the mean accuracy per tile.

it gives the possibility of handling a large amount of data without running out of memory. In this experiment, COLMAP is applied with the same configuration to each set of images and as a result, two dense point clouds are obtained. The configuration includes, apart from the default parameters, using a single camera per flight path and the vocabulary tree method [27] for feature matching. This was selected because it is the recommend mode for large image collections (several thousands). Moreover, as in COLMAP there is no option implemented to enforce GPS priors during SfM computation, we follow the recommendation of applying the geo-registration after obtaining the sparse point cloud.

The point cloud associated with the top view images contains twenty-five million points whereas the one associated with the oblique images, containing twenty-two million points, is less dense. During the process, the point clouds are coarsely registered with the GPS information to the LiDAR point cloud. The GPS information available for the top view images is more accurate than the one available for the oblique ones but a fine registration with an ICP method [2] (including scaling) is needed in both cases. The reconstructed point cloud is split into the same tiles as the LiDAR point cloud, each of which covers an area of 500x500 m^2. Overlap of these three point clouds is in twelve tiles only, which are shown in green in Figure 4. This green area is the one under study, it is numbered for referential purposes and it allows for comparison of the reconstructions in different areas of the city.

Some qualitative results are shown in Figure 5 and from a quantitative perspective, it can be observed in Figure 4 that the LiDAR point cloud is more than four times denser than the ones obtained with image-based reconstruction. A denser image-based reconstruction is obtained with the top view data in every tile. At the bottom, the same figure shows the mean accuracy from each reconstruction to the LiDAR point cloud. It shows that, on average, the oblique imagery is closer to the ground truth. This difference is approximately 2 cm in the majority of them, however, 3 of the tiles (2, 4 and 12) present a larger difference. As can be seen in Figure 4, these correspond to the tiles with a river, and they are followed by tile 8, which is also in the area.

As pointed out in [11], the mean distance between the point clouds can be affected by outliers. Hence, they propose to use the following measurements for further study: precision, recall, and F score. The precision, P, shows the accuracy of the reconstruction, the recall, R, is related to how complete the reconstruction is, and the F score, F, is a combination of both. They are defined in Eq. (1) for a given threshold distance d. In the equations, I

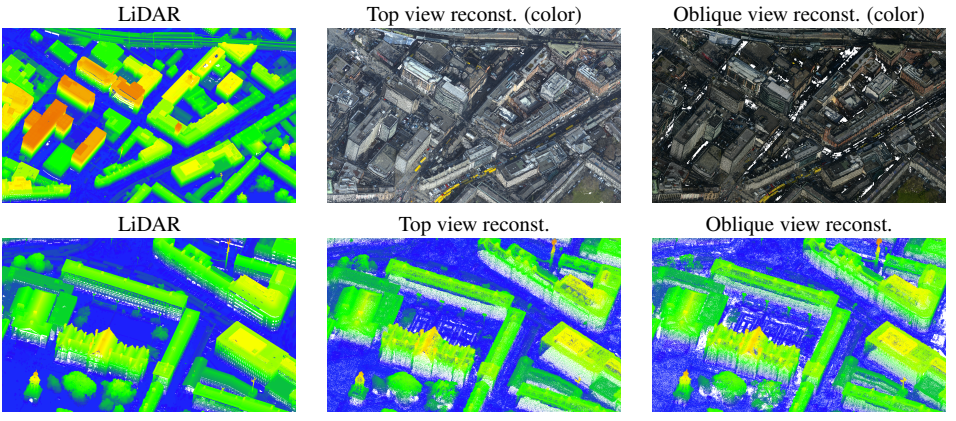


Figure 5: **Qualitative results.** Comparison of the LiDAR with the image-based 3D reconstructions in two different parts of the city. Row I: including color. Row II: only geometry.

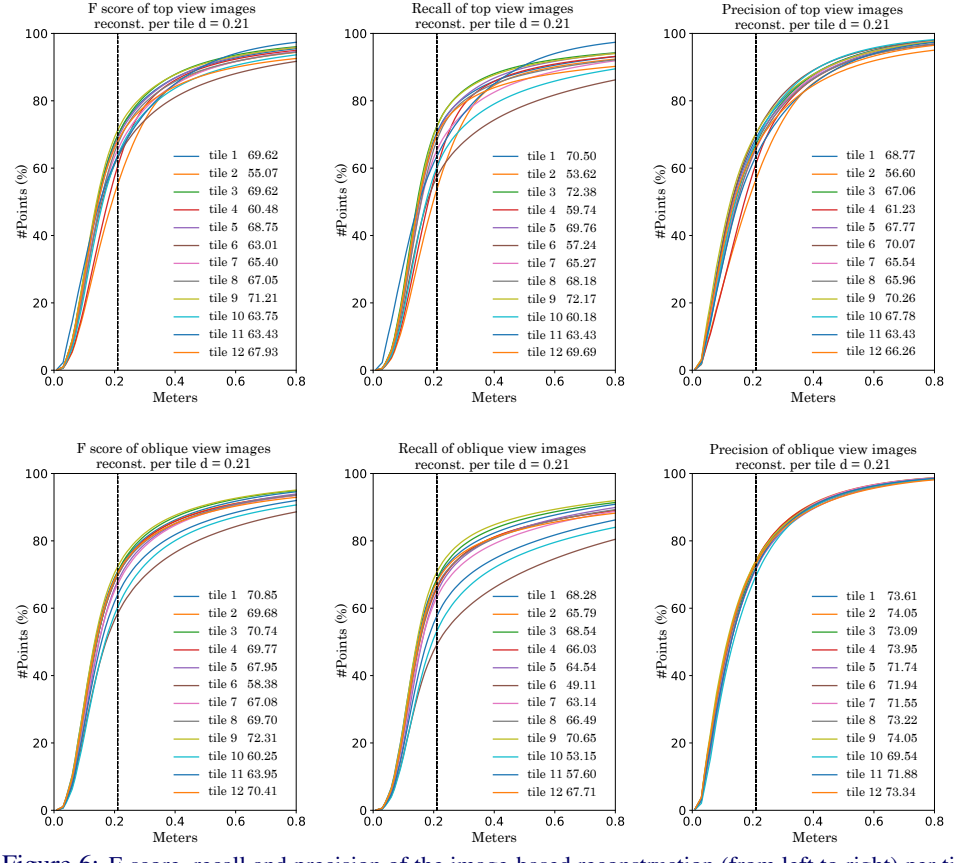


Figure 6: F-score, recall and precision of the image-based reconstruction (from left to right) per tile. The top row shows the results of the top view images reconstruction and the bottom one, the results with the oblique view ones. The values given per tile corresponds with the value of the curves at a given distance (dash line).

is the image-based reconstruction point cloud, G is the ground truth point cloud, $|\cdot|$ is the cardinality, $dist_{I\to G}(d)$ are the points in I with a distance to G less than d and $dist_{G\to I}(d)$ is analogous (*i.e.* $dist_{A\to B}(d) = \{a \in A \mid \min_{b\in B} \|a-b\|_2 < d\}$, A and B being point clouds).

$$P(d) = \frac{|dist_{I\to G}(d)|}{|I|}100 \qquad R(d) = \frac{|dist_{G\to I}(d)|}{|G|}100 \qquad F(d) = \frac{2P(d)R(d)}{P(d)+R(d)} \tag{1}$$

The results of these three measurements are given in Figure 6. From these, the results of the top view reconstruction are more tile dependant than the ones obtained with the oblique imagery. The precision in the latter is very similar for every tile, however, the recall is much lower in tiles 6, 10 and 11. A commonality amongst these tiles is that they contain part of the parks of the city, as shown in Figure 4. Apart from that, tile 9 has a higher F score, which corresponds to an area without any river or green areas.

5 Conclusions and Future Work

This paper presents a highly dense, precise and diverse labelled point cloud. Herein, an extensive manually annotated point cloud dataset is introduced for Dublin City. This work processes a LiDAR dataset that was unstructured point cloud of Dublin City Centre with various types of urban elements. The proposed benchmark point cloud dataset is manually labelled with over 260 million points comprising of 100'000 objects in 13 hierarchical multi-level classes with an average density of 348.43 points/m^2. The intensive process of labelling is precisely cross-checked with expert supervision. The performance of the proposed dataset is validated on two salient applications. Firstly, the labelled point cloud is employed for classifying 3D objects using state-of-the-art CNN based models. This task is a vital step in a scene understanding pipeline (*e.g.* urban management). Finally, the dataset is also utilised as a detailed ground truth for evaluation of image-based 3D reconstructions. The dataset will be publicly available to the community.

In addition to the above, the usage of the dataset can be extended in several applications. For example, the annotated dataset can be used for a further evaluation of image-based 3D reconstruction per class instead of per tile. Also, it can be employed for object segmentation in remote sensing or Geographic Information System (GIS), volumetric change detection for forest monitoring, as well as in disaster management. Additionally, it can be applied to optimise traffic flow for smart cities and even for the generation of real models of large cityscapes in the entertainment industry.

6 Acknowledgement

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under the Grant Number 15/RP/2776 and in part by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 780470. The authors highly appreciate the original work of generating LiDAR Point Cloud at Urban Modelling Group in University College Dublin in 2015. In addition, we are grateful for all of the volunteers who generously participated in the process of data labelling, especially Mr S Pouria Vakhshouri Kouhi for his constant support. We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

[1] AHN [Actueel Hoogtebestand Nederland dataset]. http://www.ahn.nl/common-nlm/open-data.html, 2017.

[2] Cloudcompare (version 2.10.1) [gpl software]. https://www.danielgm.net/cc/, 2019.

[3] Henrik Aanæs, Rasmus Ramsbøl Jensen, George Vogiatzis, Engin Tola, and Anders Bjorholm Dahl. Large-scale data for multiple-view stereopsis. *International Journal of Computer Vision*, 120(2):153–168, 2016.

[4] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. Patchmatch: A randomized correspondence algorithm for structural image editing. In *ACM Transactions on Graphics (ToG)*, volume 28, page 24. ACM, 2009.

[5] András Bódis-Szomorú, Hayko Riemenschneider, and Luc Van Gool. Fast, approximate piecewise-planar modeling based on sparse structure-from-motion and superpixels. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 469–476, 2014.

[6] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5828–5839, 2017.

[7] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010.

[8] Ananya Gupta, Jonathan Byrne, David Moloney, Simon Watson, and Hujun Yin. Automatic tree annotation in lidar data. In *Proceedings of the International Conference on Geographical Information Systems Theory, Applications and Management*, pages 36–41, 2018.

[9] Timo Hackel, Nikolay Savinov, Lubor Ladicky, Jan D. Wegner, Konrad Schindler, and Marc Pollefeys. Semantic3d.net: A new large-scale point cloud classification benchmark. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1/W1:91–98, 2017. doi: 10.5194/isprs-annals-IV-1-W1-91-2017. URL https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/IV-1-W1/91/2017/.

[10] Latika Kharb. Innovations to create a digital india: Distinguishing reality from virtuality. *Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org*, 6(9), 2016.

[11] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)*, 36(4):78, 2017.

[12] Debra F Laefer, Saleh Abuwarda, Anh-Vu Vo, Linh Truong-Hong, and Hamid Gharibi. 2015 aerial laser and photogrammetry survey of dublin city collection record. https://geo.nyu.edu/catalog/nyu-2451-38684.

[13] Stefan Lang, Thomas Blaschke, Gyula Kothencz, and Daniel Hölbling. 13 urban green mapping and valuation. *Urban Remote Sensing*, page 287, 2018.

[14] Jiaxin Li, Ben M. Chen, and Gim Hee Lee. So-net: Self-organizing network for point cloud analysis. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[15] Davide Marelli, Simone Bianco, Luigi Celona, and Gianluigi Ciocca. A blender plug-in for comparing structure from motion pipelines. In *IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, pages 1–5, 2018.

[16] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 922–928. IEEE, 2015.

[17] Xuelian Meng, Nate Currit, and Kaiguang Zhao. Ground filtering algorithms for airborne lidar data: A review of critical issues. *Remote Sensing*, 2(3):833–860, 2010.

[18] Pierre Moulon, Pascal Monasse, and Renaud Marlet. Global fusion of relative motions for robust, accurate and scalable structure from motion. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 3248–3255, 2013.

[19] F Pirotti. Analysis of full-waveform lidar data for forestry applications: a review of investigations and methods. *iForest-Biogeosciences and Forestry*, 4(3):100, 2011.

[20] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 652–660, 2017.

[21] Charles R. Qi, Wei Liu, Chenxia Wu, Hao Su, and Leonidas J. Guibas. Frustum pointnets for 3d object detection from rgb-d data. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[22] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*, pages 5099–5108, 2017.

[23] Aakanksha Rana, Cagri Ozcinar, and Aljosa Smolic. Towards generating ambisonics using audio-visual cue for virtual reality. In *44th International Conference on Acoustics, Speech, and Signal Processing, (ICASSP)*, 2019.

[24] Aakanksha Rana, Giuseppe Valenzise, and Frederic Dufaux. Learning-based tone mapping operator for efficient image matching. *IEEE Transactions on Multimedia*, 21(1):256–268, Jan 2019.

[25] Xavier Roynard, Jean-Emmanuel Deschaud, and François Goulette. Paris-lille-3d: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. *The International Journal of Robotics Research*, 37(6):545–557, 2018.

[26] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[27] Johannes Lutz Schönberger, True Price, Torsten Sattler, Jan-Michael Frahm, and Marc Pollefeys. A vote-and-verify strategy for fast spatial verification in image retrieval. In *Asian Conference on Computer Vision (ACCV)*. Springer, 2016.

[28] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*. Springer, 2016.

[29] Thomas Schops, Johannes L Schonberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, and Andreas Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3260–3269, 2017.

[30] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 519–528, 2006.

[31] Noah Snavely, Steven M Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, 2008.

[32] Christoph Strecha, Wolfgang Von Hansen, Luc Van Gool, Pascal Fua, and Ulrich Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.

[33] Christopher Sweeney, Tobias Hollerer, and Matthew Turk. Theia: A fast and scalable structure-from-motion library. In *Proceedings of the ACM International Conference on Multimedia*, pages 693–696. ACM, 2015.

[34] Yuliya Tarabalka and Aakanksha Rana. Graph-cut-based model for spectral-spatial classification of hyperspectral images. In *2014 IEEE Geoscience and Remote Sensing Symposium*, pages 3418–3421, July 2014.

[35] Bruno Vallet, Mathieu Brédif, Andrés Serna, Beatriz Marcotegui, and Nicolas Paparoditis. Terramobilita/iqmulus urban point cloud analysis benchmark. *Computers & Graphics*, 49:126–133, 2015.

[36] Shenlong Wang, Min Bai, Gellert Mattyus, Hang Chu, Wenjie Luo, Bin Yang, Justin Liang, Joel Cheverie, Sanja Fidler, and Raquel Urtasun. Torontocity: Seeing the world with a million eyes. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 3028–3036, 2017.

[37] Andreas Wichmann, Amgad Agoub, and Martin Kada. Roofn3d: Deep learning training data for 3d building reconstruction. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(2), 2018.

[38] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3D shapenets: A deep representation for volumetric shapes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1912–1920, 2015.

[39] Liqiang Zhang, Zhuqiang Li, Anjian Li, and Fangyu Liu. Large-scale urban point cloud labeling and reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 138:86–100, 2018.

[40] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4490–4499, 2018.