This is the preprint version of the paper:
Xue, F., Chen, K. & Lu W. (2019). Understanding unstructured 3D point clouds for creating digital twin city: An unsupervised hierarchical clustering approach. *CIB World Building Congress 2019*, 17-21 June, 2019, Hong Kong SAR.

Understanding unstructured 3D point clouds for creating digital twin city: An unsupervised hierarchical clustering approach

Fan Xue,
Department of Real Estate and Construction
The University of Hong Kong
Hong Kong SAR
(email: xuef@hku.hk)
Ke Chen,
Department of Construction Management
Huazhong University of Science and Technology
Hubei, China PR
(email: chenkecm@hust.edu.cn)
Weisheng Lu,
Department of Real Estate and Construction
The University of Hong Kong
Hong Kong SAR
(email: wilsonlu@hku.hk)

Abstract

Digital twin city (DTC) is a critical information infrastructure that enables many innovative applications for smart and resilient city development. Thanks to the recent advances in remote sensing and photogrammetry, accurate, dense, and large-scale 3D urban point clouds become increasingly available for many cities for creating and updating their DTCs. Because of the immense amount and the high update frequency of urban point clouds, it is too time-consuming and labor-intensive to create and update DTCs solely by human experts. Researchers have developed a wealth of automatic and semi-automatic methods for processing 3D urban point clouds using expert knowledge of the built environment, supervised learning, and reinforced learning of geometric primitives and components. However, these methods are restricted, ironically by the embedded knowledge, in the scalability to sophisticated scenes and the availability of standardized components.

Inspired by the success of Google' unsupervised learning program AlphaZero, this paper proposes a novel hierarchical clustering approach for semantic enrichment of point clouds. Unlike the existing approaches relying on fixed domain knowledge, extra correlational training examples, or available 3D references, the proposed approach exploits the similarities between patches of point clouds without explicit domain knowledge. The proposed approach first segments patches from the input point cloud through the connected subgraphs of voxel grids, then computes the dissimilarity matrix between the patches via iterative optimization. Subsequently, the dissimilarity engenders a hierarchy of clusters for understanding the relatedness between the patches. A pilot study on a real urban scene showed that the proposed approach is feasible and potent to cluster and detect objects automatically. Another experiment showed that the dissimilarity-based clusters and associated transformations can help create semantic objects for DTC, as referential 3D models are available.

Keywords: Digital twin city (DTC), LIDAR point cloud, hierarchical clustering, object detection and scene understanding, urban semantics

1. Introduction

A digital twin is a virtual representation of a physical object or system across its lifecycle, using real-time data to enable understanding, learning, and reasoning (NIC 2017). Based on its digital twin, a complex object such as aircraft or factory production line can be monitored without close proximity to the physical object. Furthermore, analysis and simulations of digital twins can help reveal and mitigate unpredictable and undesirable emergent behaviors of the complex objects (Grieves & Vickers 2017).

A digital twin city (DTC), likewise, is a virtual representation of the lifecycles of physical objects and assets in a city. Buildings, road networks, vegetation, automobiles, as well as residents in a city do not pop into the built environment, but progress through a lifecycle of planning, creation (e.g., procreation and migration of residents), operation, and disposal (or cessation of residence). Singapore has perhaps the world first DTC which was entitled "Virtual Singapore" and collaborated with *Dassault Systèmes* (NRF 2018). DTC that holds the lifecycle information, by definition, is a superset of many computer-aided technologies including building information modeling (BIM) (NIBS 2015; Sacks et al. 2018) and geographic information system (GIS) (Burrough et al. 2015). DTC is thus a critical information infrastructure enabling many innovative applications for smart and resilient city development (Kitchin 2014), including urban planning, architectural design, construction management, human and natural geography, transportation and accessibility, resource conservation, and robotics and self-driving cars.

Spatiotemporal information is the most fundamental in DTC like in BIM (Xue et al. 2018). Recent advances in laser and vision-based remote sensing technologies, e.g., laser scanning and photogrammetry, have led to accurate, dense, large-scale, and – most importantly – affordable spatiotemporal data, such as 3D point clouds, of many cities, e.g., Light Detection and Ranging (LiDAR) data of Hong Kong (CEDD 2015) and Dublin, Ireland (Laefer et al. 2017). Based on the spatiotemporal data, more non-geometric semantics including category, materials, functions, accessing instructions, and topological relationships can be enriched for creating and updating the DTC (Xue et al. 2019a; Xue et al. 2019b). Because of the immense amount and the high update frequency of urban objects, it is too tedious, time-consuming, and costly to create and update digital urban objects solely by human experts (Xue et al. 2018). On the other hand, current automatic and semi-automatic methods are limited by (i) the complexity of urban systems, (ii) the scalability to sophisticated scenes, or (iii) the availability of standard 3D referential components.

Inspired by the success of Google's unsupervised learning program AlphaZero in multiple games including Go, chess (28 wins, 72 draws and 0 losses against Stockfish), and shogi (Silver et al. 2018), this paper explores a novel, automatic, unsupervised hierarchical clustering approach for creating DTC. The proposed approach first clusters the 3D point clouds of unknown urban objects in a city through geometric similarity, then understands the hierarchical relationships between the unknown objects – as well as known standard models. Pilot experiments on a real-world scene in Dublin, Ireland was conducted to validate the approach. The contribution of this paper is two-fold. First, the proposed approach extends the methodological boundary of semantic enrichment to unsupervised machine learning techniques for point cloud data in theory. Secondly, the implementation based on an open source software library and real data confirmed the feasibility and potential of industrial applications.

2. Related work

A wealth of automatic and semi-automatic semantic enrichment methods have been developed for correlating new information to 3D measurement data like point clouds for creating DTC (Huber et al. 2011; Wang & Kim 2019). In a machine learning perspective, all the methods can be classified into three groups based on the behaviors of their inner correlation models. The first group is "no learning," in which the correlations are made of structured expert knowledge (e.g., geometric primitives and basic rules) on the built environment, e.g., (Huber et al. 2011; Valero et al. 2012; Zou et al. 2018). The second group is "supervised learning" (e.g., semantic segmentation), where a supervised training is performed to correlate the manual labels to the measurement features in training examples, e.g., (Xiong et al. 2013; Babacan et al. 2017; Czerniawski et al. 2018; Wang et al. 2018). The third group is "reinforced learning" (e.g., semantic registration), in which the correlation was gradually converged after iterations of trials of possible targets (e.g., online open 3D components), e.g., (Xue et al. 2018; Xue et al. 2019a; Hidaka et al. 2018).

However, the first group of methods is mainly restricted to the regularity in pre-assumed scenes (e.g., a box-shaped indoor space) (Valero et al. 2012; Zou et al. 2018). The supervised learning itself, in the second group, is very challenging in sophisticated urban scenes (Babacan et al. 2017); in addition, preparation of the training labels also costs a fortune. The third group is limited to the availability of standardized 3D components for matching the input points (Xue et al. 2019a). In summary, each group of existing methods has its drawbacks.

Unsupervised learning, which is a parallel stream to supervised learning and reinforced learning, requires neither training examples nor standard components. Clustering, the most common unsupervised learning method, groups a set of objects so that the objects in one group (cluster) are more similar than those in other groups. Examples of clustering methods for point clouds are the region growing (Pauly et al. 2002), *k*-means (Shi et al. 2011), and supervoxel clustering (Papon et al. 2013). As far as is concerned, most unsupervised learning studies stopped at patch (part of cloud) segmentation and failed to involve the urban semantics in the clouds for semantic enrichment and creation of DTCs.

3. The proposed method for semantic enrichment

This study investigates an unsupervised hierarchical clustering approach for semantic enrichment of 3D point clouds. As shown in *Figure 1*, the only input to the approach is a cloud of unstructured points. First, the input point cloud is preprocessed to remove reference surfaces (e.g., ground point removal) to form a disconnected point cloud. Then, the supervoxels and their connectivity are analyzed using the supervoxel clustering (Papon et al. 2013), of which the codes are from the open source software library *point cloud library* (pcl, version 1.8). The connected graphs of the voxel grids are clustered as patches, as shown in *Figure 1*. The patches are then compared to compute the dissimilarity as the minimum root-mean-square error (RMSE) after every possible rotation and translation:

$$dissimilarity(P_i, P_j) = \min_{r,\, t \in \mathbb{R}^3} RMSE(P_i, \text{translate}(\text{rotate}(P_j, r), t)), \quad (1)$$

where the *RMSE* function measure the error between the points in two patches, $r = [r_x, r_y, r_z]$ is the 3D rotation parameters defined on $\mathbb{R}^3$, and $t = [t_x, t_y, t_z]$ is the 3D translation parameters. We adopt an open source library ODAS (Optimization-based Detection of Architectural Systems) (Xue et al. 2019b) and rewrote the objective function to *Equation 1*, to compute the dissimilarity matrix between all the patches. Finally, the dissimilarity matrix produces a tree of hierarchical clusters via the *scipy* package (ver. 0.19, in Python), where the similar patches are listed together. A threshold, e.g., 30% of maximum dissimilarity or fixed value of 10 cm, can cut the tree into clusters (groups). The connected patches and their clusters are the output of the approach, can be enriched to the semantics of input point cloud, e.g., as user-defined properties of "group," "subgroup," and "most similar object."

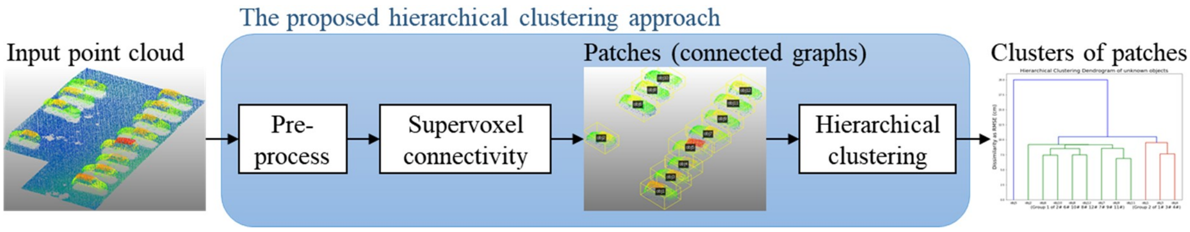
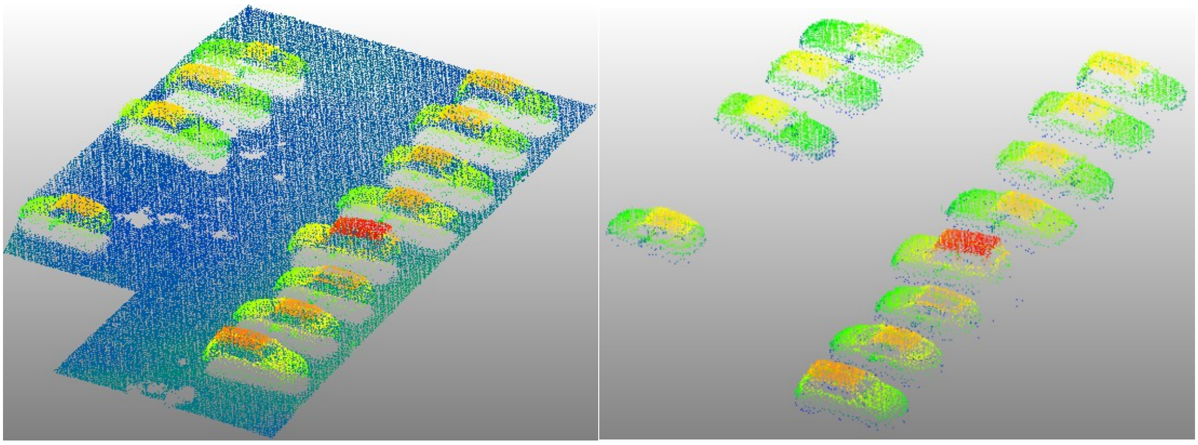


Figure 1: An overview of the proposed hierarchical clustering approach

4. Experimental tests on a pilot case

4.1 A pilot case

A pilot study was conducted on a case of small-scale LiDAR point cloud scanned from Dublin, Ireland (Laefer et al. 2017). The selected cloud, consisting of 112,999 points (6.78MB compressed on disk) as shown in *Figure 2*, was a car park scene with 12 city cars. The cars included 8 "short" cars (height < 1.5m), 3 "tall" cars (1.5m ≤ height < 1.9m), 1 full-size SUV (sport-utility vehicle, height ≥ 1.9m) of various models, parking locations, and orientations. A preprocess of planar removal removed almost all points of the ground. The cloud after the ground removal consisted of only 24,126 points.

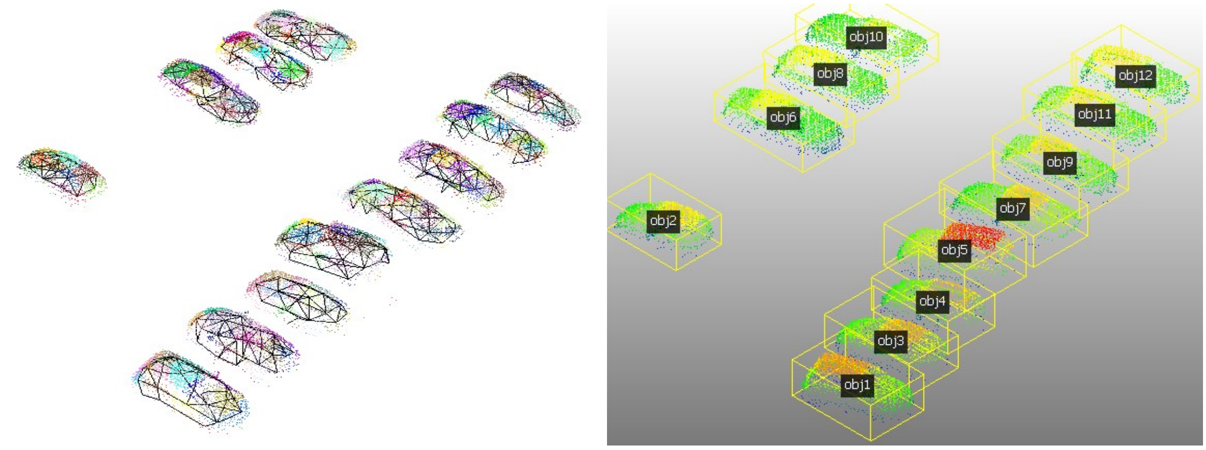


(a) 112,999 LiDAR points (color indicates height) (b) After the preprocess of planar removal

Figure 2: The pilot case of a car park scene

4.2 Experimental results

Then, we applied the supervoxel clustering algorithm, with the parameter "voxel seed size = 15cm." The results were obtained in 1.3 seconds, including 368 small patches and the connectivity among them, as shown in *Figure 3*a. Based on the connected subgraphs, 12 patches were formed and named from obj_1 to obj_{12} based on the centroids along the x-axis, as shown in *Figure 3*b.

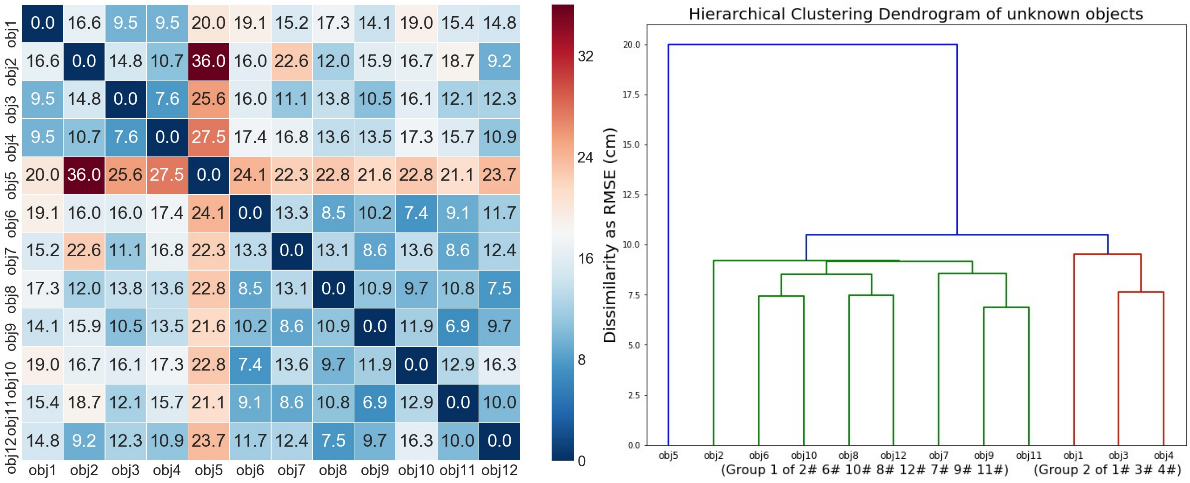


(a) 368 small patches (by color) and the connectivity (lines) detected in 1.3s

(b) 12 patches (obj_1 to obj_{12}) was clustered via the connectivity of patches in (a)

Figure 3: The supervoxel connectivity and patch (connected graphs) clustering

The next step is the hierarchical clustering. First, the dissimilarity matrix was calculated between all 12 patches in 109.4s, where the parameters were set to the default values in Xue et al. (2019b). The matrix is visualized in *Figure 4*, where it can be found that the obj_5 seems much different from other patches. Meanwhile, a tree of similar patches was created from the matrix. As shown in *Figure 4*, the obj_5, which was an SUV, was first distinguished from the rest patches; while the remaining patches were similar. By applying a grouping threshold of 10cm, we got two groups. One group, in red in *Figure 4*, included three patches of obj_1, obj_3, and obj_4, which were all the "tall" cars (with roof height at 1.5~1.6m). The other group in green was the set of "short" cars (with height at 1.3~1.4m). For each point shown in *Figure 3*b, the name of the clustered patches and the group of patches were added to new properties and saved in the Stanford Polygon (.ply) format. Therefore, the results confirmed the geometric dissimilarity and unsupervised clusters detected by the proposed hierarchical clustering approach.

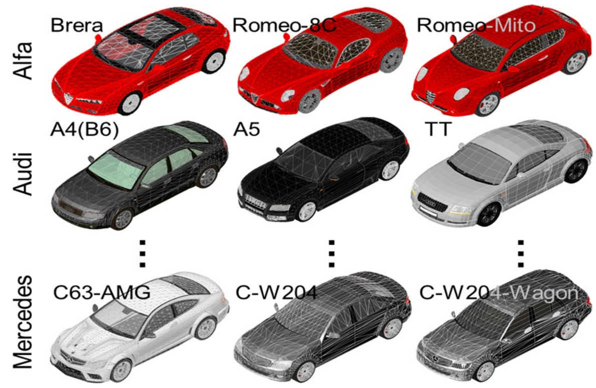


(a) The dissimilarity matrix computed in 109.4s (unit = cm, color depth indicates the dissimilarity) (b) Hierarchical clusters of similar patches (grouping threshold = 10cm)

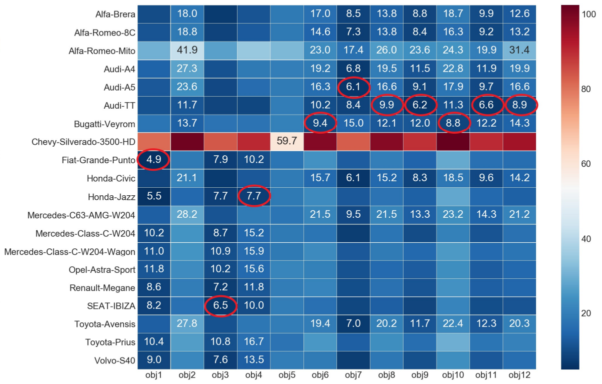
Figure 4: Results of the proposed hierarchical clustering approach

4.3 Application to unsupervised segmentation for digital twin city

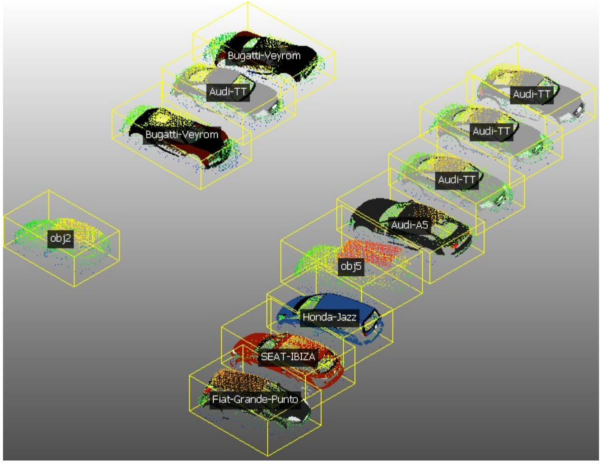
The semantically enriched point cloud can be applied to unsupervised point cloud segmentation for creating DTC. For example, we downloaded the CAD (computer-aided design) models of 20 different city car models from the manufacture's and fan's websites, as shown in *Figure 5*a. The models included popular brands such as Alfa, Audi, Honda, Toyota, Mercedes-Benz, and Volvo. The "reinforced learning" semantic registration methods used to consume hours to optimize the location and direction of each possible model to the whole scene (Xue et al. 2019a). Now, only the error matrix (see *Equation 1*) between the 12 patches and the 20 models, i.e., 12 × 20 = 240 pairwise comparisons, is needed for assigning each patch to a car model. Furthermore, the CAD models can also be grouped using the proposed hierarchical clustering approach, such that the 240 comparisons were reduced to 108 with a focus on the groups in similar sizes. The 108 comparisons were computed in 152.8s as shown in *Figure 5*b, and the most similar car models were selected according to the minimum dissimilarity. It can be found in *Figure 5*b that the minimum dissimilarity (in darker blue) for each patch was included in the 108 pairs, i.e., saving more than 50% time without loss on the optimal solutions. Also, the obj_2 (a subcompact) and obj_5 (an SUV) did not match very similar cars (error threshold = 10cm) in the 20 given models. The selected city car models, along with rich semantics such as class, brand, model, production, and performance, were then registered to the patches for creating a digital city twin, as shown in *Figure 5*c.



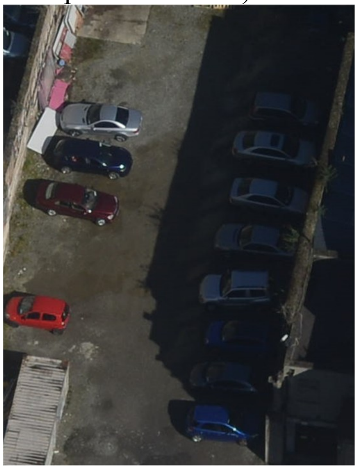
(a) Online open CAD files of 20 known car models



(b) The dissimilarity matrix (unit = cm, best model for each patch is circled)



(c) Registered car models for a digital twin



(d) Ground truth

Figure 5: Results of registration standard car CAD models to the clustered patches

*Figure 5*d shows the ground truth about the 12 cars in an aerial photo. It can be found that the hierarchical clusters in *Figure 4* were meaningful and correct. Although some CAD models in *Figure 5*c, e.g., the two Bugatti Veyrons, were imperfect in terms of geometry, the locations and heading directions were correct in general. In summary, the proposed unsupervised hierarchical clustering approach can resolve a critical issue, i.e., the scalability to large-scale point clouds, for other semantic enrichment methods such as the "reinforced learning" semantic registration.

5. Discussion and conclusion

In the era of smart cities, the urban semantics embedded in real-time digital twin city (DTC) is vital for many innovative applications. Yet, the task of enriching the measurement data such as 3D point clouds is very challenging. This paper presents an unsupervised hierarchical clustering approach for the semantic enrichment of point cloud, for facilitating the creation of DTC. The approach first forms patches as the connected subgraphs of supervoxel regions, then clusters the patches based on the dissimilarity matrix between them. In contrast to the existing methods, the proposed approach enriches object-level semantics to the input point cloud without prerequisites of correlational training labels or available 3D referential models. A pilot test on a car park scene in Dublin, Ireland confirmed the feasibility and meaningfulness of the proposed hierarchical clustering approach. The output of the approach can facilitate the creation of DTC, such as the unsupervised point cloud segmentation. Future work includes (i) triangulation of urban regularity (e.g., symmetry and repetition of the urban objects), (ii) automatic selection and adaptation of algorithms and parameters for various scenes, and (iii) integration to existing data standards and software related to DTC.

Acknowledgements

The authors would like to acknowledge the support by the Hong Kong Research Grant Council, Grant Nos. 17201717 and 17200218, and The University of Hong Kong, Grant Nos. 201811159177 and 102009741.

References

Babacan, K., Chen, L. & Sohn, G. (2017). Semantic segmentation of indoor point clouds using Convolutional Neural Network. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-4*(W4), 101-108.

Burrough, P. A., McDonnell, R., McDonnell, R. A. & Lloyd, C. D. (2015). *Principles of geographical information systems.* Oxford university press.

CEDD. (2015). *The CEDD 2010 LiDAR Survey (private communication).* Hong Kong: Civil Engineering and Development Department.

Czerniawski, T., Sankaran, B., Nahangi, M., Haas, C. & Leite, F. (2018). 6D DBSCAN-based segmentation of building point clouds for planar object classification. *Automation in Construction, 88*, 44-58. doi:10.1016/j.autcon.2017.12.029

Grieves, M. & Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In F.-J. Kahlen, S. Flumerfelt & A. Alves, *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches* (pp. 85-113). Springer.

Hidaka, N., Michikawa, T., Motamedi, A., Yabuki, N. & Fukuda, T. (2018). Polygonization of point clouds of repetitive components in civil infrastructure based on geometric similarities. *Automation in Construction, 86*, 99-117. doi:10.1016/j.autcon.2017.10.014

Huber, D., Akinci, B., Oliver, A. A., Anil, E., Okorn, B. E. & Xiong, X. (2011). Methods for automatically modeling and representing as-built building information models. *Proceedings of the NSF CMMI Research Innovation Conference.* Retrieved September 18, 2018, from https://ri.cmu.edu/pub_files/2011/1/2011-huber-cmmi-nsf-v4.pdf

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal, 79*(1), 1-14. doi:10.1007/S10708-013-9516-8

Laefer, D. F., Abuwarda, S., Vo, A.-V., Truong-Hong, L. & Gharibi, H. (2017). *2015 Aerial Laser and Photogrammetry Survey of Dublin City Collection Record*. doi:10.17609/N8MQ0N

NIBS. (2015). *National Building Information Modeling Standard (Version 3).* National Institute of Building Sciences. Retrieved from https://www.nationalbimstandard.org/

NIC. (2017). *Data for the Public Good.* London: National Infrastructure Commision, UK. Retrieved from https://www.nic.org.uk/publications/data-public-good/

NRF. (2018). *Virtual Singapore.* Singapore: National Research Fundation. Retrieved from https://www.nrf.gov.sg/programmes/virtual-singapore

Papon, J., Abramov, A., Schoeler, M. & Worgotter, F. (2013). Voxel cloud connectivity segmentation-supervoxels for point clouds. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2027-2034). IEEE.

Pauly, M., Gross, M. & Kobbelt, L. P. (2002). Efficient simplification of point-sampled surfaces. *Proceedings of the IEEE conference on Visualization'02* (pp. 163-170). IEEE.

doi:10.1109/VISUAL.2002.1183771

Sacks, R., Eastman, C. M., Lee, G. & Teicholz, P. (2018). *BIM Handbook: A guide to Building Information modeling for owners, designers, engineers, contractors, and facility managers* (3rd ed.). Hoboken, NJ, USA: John Wiley & Sons.

Shi, B. Q., Liang, J. & Liu, Q. (2011). Adaptive simplification of point cloud using k-means clustering. *Computer-Aided Design, 43*(8), 910-922. doi:10.1016/j.cad.2011.04.001

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T. & Lillicrap, T. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science, 362*(6419), 1140-1144. doi:10.1126/science.aar6404

Valero, E., Adán, A. & Cerrada, C. (2012). Automatic method for building indoor boundary models from dense point clouds collected by laser scanners. *Sensors, 12*(12), 16099-16115. doi:10.3390/s121216099

Wang, Q. & Kim, M. K. (2019). Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018. *Advanced Engineering Informatics, 39*, 306-319. doi:10.1016/j.aei.2019.02.007

Wang, W., Yu, R., Huang, Q. & Neumann, U. (2018). SGPN: Similarity group proposal network for 3D point cloud instance segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2569-2578). IEEE.

Xiong, X., Adan, A., Akinci, B. & Huber, D. (2013). Automatic creation of semantically rich 3D building models from laser scanner data. *Automation in Construction, 31*, 325-337. doi:10.1016/j.autcon.2012.10.006

Xue, F., Lu, W. & Chen, K. (2018). Automatic generation of semantically rich as-built building information models using 2D images: A derivative-free optimization approach. *Computer-Aided Civil and Infrastructure Engineering, 33*(11), 926-942.

Xue, F., Lu, W., Chen, K. & Zetkulic, A. (2019a). From 'semantic segmentation' to 'semantic registration': A derivative-free optimization-based approach for automatic generation of semantically rich as-built building information models (BIMs) from 3D point clouds. *Journal of Computing in Civil Engineering*, in press.

Xue, F., Lu, W., Webster, C. & Chen, K. (2019b). A derivative-free optimization-based approach for detecting architectural symmetries from 3D point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing, 148*, 32-40.

Zou, C., Colburn, A., Shan, Q. & Hoiem, D. (2018). LayoutNet: Reconstructing the 3D room layout from a single RGB image. *2018 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2051-2059). Salt Lake City, USA: IEEE.