ISSN (Print): 2476-8316
ISSN (Online): 2635-3490
Dutse Journal of Pure and Applied Sciences (DUJOPAS), *Vol. 4 No. 2 December 2018*

3D Sensing and Learning: A Review

[1,2]**Jibril Muhammad Adam,** [1]**Aminu Abdullahi Aliyu,**
[1]**Muhammad Abubakar Alhassan**
[1]Department of Computer Science
Federal University Dutse
[2]Fujian Key Laboratory on Sensing and Computing for Smart Cities
Xiamen University

Abstract

3D data provides more information about the world we live in and with the state-of-the-art results obtained from 2D vision tasks, attention has been shifted to extending these processes to 3D. The availability of 3D datasets, advancements in hardware and resurgence of neural networks (deep learning) has made this extension possible albeit more challenging. Encouraging and improving results are emerging in 3D processing tasks like classification, segmentation and recognition. This paper provides an overview of the different 3D sensing methods and representations. It also contains a brief review of deep learning (DL) architectures and frameworks classified based on input format.

Keywords: 3D scanning, 3D representation, 3D processing, 3D vision tasks, CNN

INTRODUCTION

3D (three dimensions or three-dimensional) data provides the dimension of depth (z-axis) to 2D (x and y-axis) data. This third dimension makes it possible to depict the real world more accurately because we live in a three-dimensional reality. Recently, there have been great advancements in 3D data acquisition. Collection of 3D data is becoming more feasible and affordable with the emergence of low-cost and easy-to-use devices like Microsoft Kinect(Microsoft, 2018). The type of scanning device used determines the form of data produced as output. For example, LIDAR ('Laser Imaging Detection And Ranging' or 'LIght and RaDAR ')scanners produce as output 3D point cloud model of scanned scenes and devices like Microsoft Kinect produce color and depth images (RGB-D: Red, Green, Blue - Depth) which can be used for different purposes.

2D tasks on images have seen tremendous progress with some state-of-the-art results obtained in tasks such as object classification, segmentation, object recognition, and detection etc. These all started with the breakthrough result obtained by (Krizhevsky, Sutskever, & Hinton, 2012) in ImageNet Large Scale Visual Recognition Competition, 2012 (ILSVRC). This revolutionalized the field of computer vision and set the tone for some of the outstanding results reported on 2D tasks. The increasing availability of 3D data, hardware improvement (GPU and TPU), advances in machine and deep learning and the potential benefits that will be gained from successful 3D tasks makes reviewing 3D sensing, learning, and perception very important and necessary. 3D learning and perception has a lot of potential in the field of robotics, augmented and virtual

*Author for Correspondence

reality, medical imaging, autonomous driving etc.(Ioannidou, Chatzilari, Nikolopoulos, & Kompatsiaris, 2017).

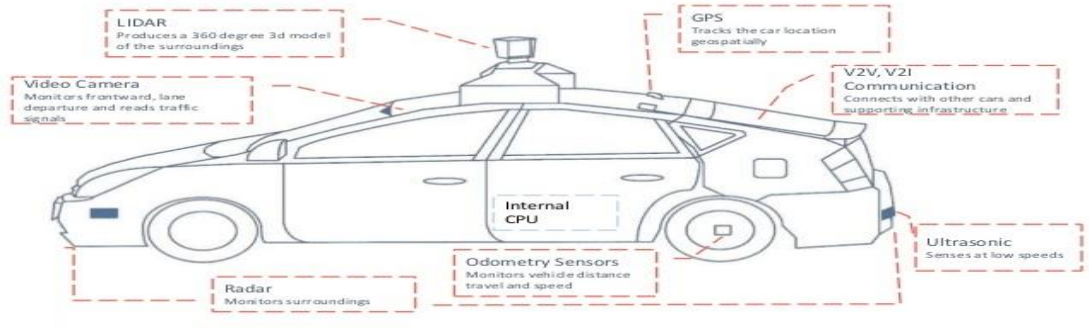

Figure 1: Autonomous vehicle technology (https://www.todaysoftmag.ro/article/2696/)

The methodology employed in this paper is the same as in (Ahmed, et al., 2018)where 3D data representations and format provide the basis for the study, categorization and understanding of 3D vision tasks and architectures.The methodology used in Section II uses the technology of range imaging and depth sensors as the basis for the review and understanding of different sensing technologies, same approach is used (Mahony, Campbell, Krpalkova, & Riordan, 2018)and (Schoning & Heidemann, 2016). The scope restricts the review to the areas covered by the methodology.

This paper starts by presenting the different ways of data acquisition (sensing) in 3D (Section II). Section III looks at the different 3D representations and Section IV tackles the concept of learning in 3D. 3D architectures and trends are discussed in Section V and Section VI provides future and promising research areas concluded by final remarks.

SENSING IN 3D

Sensing in 3D is more complicated and expensive than in 2D, but the emergence of Microsoft Kinect in November 2010 as the first consumer RGB-D camera was thought to be the beginning of low-cost and easy to use scanning devices in the market and a possible alternative to the hitherto specialized and expensive scanners e.g. laser scanners. This has not happened yet but a combination of both categories of devices is used. Sensing is generally done by capturing the target scene with a scanner, merging the information according to sensing methodology used and finally recreating a 3D model of the scene (Zhang, Dong, & El Saddik, 2015).

There are many surveys carried out on 3D sensing, sensors and techniques involved. Examples are: (Blais, 2004) surveyed off-the-shelf 3D sensors that have been around for years due to their resilience; application of 3D sensing techniques in fighting crime, medical imaging and architectural and cultural heritage preservation (Sansoni, Trebeschi, & Docchio, 2009); survey of technologies and methodologies(Daneshmand, et al., 2018), taxonomy for the assessment of 3D sensors (Schoning & Heidemann, 2016) etc. However, these papers did not address the concept of learning on 3D data and with the recent emergence of 3D datasets, upgrade in processing power and improvement in machine and deep learning algorithms, it is of paramount importance to look at learning in 3D. Sensing can be divided into parts according to different

principles e.g. depth measurement (Structured Light – SL and Time of Flight - ToF) (Schoning & Heidemann, 2016), sensing processes (Zhang, Dong, & El Saddik, 2015) and application domain (Schoning & Heidemann, 2016). In this paper, we adopt the methodology used in (Schoning & Heidemann, 2016) and (Mahony, Campbell, Krpalkova, & Riordan, 2018):

1. LIDAR: an optical remote-sensing technique that involves emitting laser light on an object and measuring how long it took to return to the sensor. This produces a densely sampled model of the object in 3D (accurate x, y, z measurements). The output model is in the form of points (thousands or millions). An example is a point cloud model of an object.

2. Digital Photogrammetry and/or Stereo: both terms involve recovering depth information (z axis) or distance of an object from multiple images or cameras. Output can be both in 2D or 3D.

3. Structured Light or Infrared Scanning: RGB-D cameras are used for this. It uses the concept of ToF or "structured light" by recreating the structure of an object due to distortions formed from projected light and camera.

The table below gives the benefits and drawbacks of each category

Table 1: Advantages and disadvantages of 3D sensing methods

Category	Advantage	Disadvantage
LIDAR	• Longer range • Captures many points • Highest quality and more robust to interference	• Most expensive • Almost always requires a specialist and a dedicated/proprietary software/scanning device
Digital Photogrammetry and/or Stereo	• Easily affordable • Requires ordinary cameras which are easier to operate	• Not suitable in situations where accuracy and speed are important • It is computationally intensive because of the matching of corresponding points in images • Textureless environments are also a bottleneck
Structured Light or Infrared Scanning	• Sits between cost and quality of scan • Fast and robust against matching errors	• More expensive cameras especially if higher quality is required • Pattern sensing failures • Blockage of objects by those in front of them causes many holes in the output • Range issues due to increase in distance between object and scanning device

3D REPRESENTATION

After capturing the data, it needs to be represented in a format that can be fed as input to an algorithm or architecture that will process it so as to produce a meaningful result. These are the four main representations of 3D data.

a. Point cloud: LIDAR and RGB-D data are captured in this format. It is a collection of data points in 3D space. It is accompanied by optional attributes like intensity, colour, flight line, time etc. Many tasks can be done using point cloud data as input, examples are classification and segmentation (Nguyen & Le, 2013).

b. Volumetric/Voxel grid: A voxel is the smallest discernible element in a 3D object.Voxels can serve as the 3D counterpart of pixels. They do not contain positional information but their positions are determined in relation to other neighboring voxels. It can also have other properties like texture and colour. Examples of tasks carried out on voxels are segmentation (Aijazi, Checchin, & Trassoudaine, 2013) and classification (Wu, et al., 2015).

c. Polygon meshes: a representation of 3D object using edges and vertices. The basic building block is a polygon shape (triangles are mostly used) and these shapes are connected together to form an object. Polygon meshes are usually used for 3D modeling and animation. 3D tasks can also be applied to this type of data format e.g. classification and segmentation (Zhu, Shen, Gao, & Hu, 2018).

d. Multi-view: 3D objects created from a collection of 2D images rendered from different viewpoints of the images. Multi-view images are usually used in transforming 3D data into a format that can be easily processed by 3D algorithms. Final output representation of multi-view objects can, therefore, be in the following four formats: depth map, point cloud, volume scalar field and mesh (Furukawa & Hernandez, 2015).

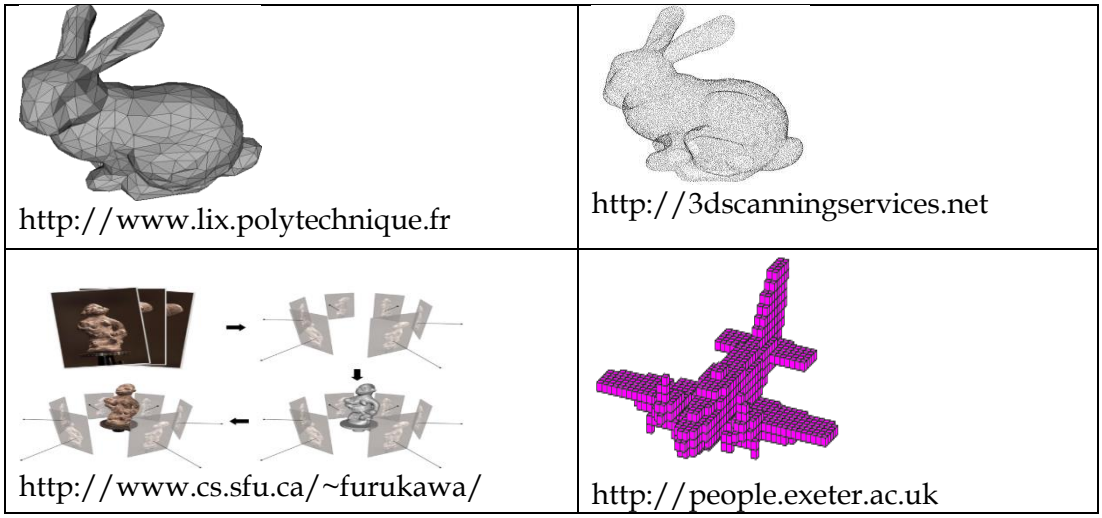

Figure 2: 3D Mesh (top left), Point cloud (top right), multi-view pipeline (bottom left), voxel grid image (bottom right)

LEARNING IN 3D

Computer vision is a subfield of artificial intelligence that deals with automation of human visual system in computers. It tries to give computers ability to see and make sense of or perceive images and videos the way humans do. The field has been around for a while and has seen many breakthroughs and disappointments as well but the recent improvements in algorithms/architectures, advancements in computer hardware and increasingly available datasets makes deep learning suitable for carrying out vision tasks. In this paper, we will focus on the architectures in the field of 3D vision. (Parvat, Chavan, Kadam, Dev, & Pathak, 2017) have reviewed the major frameworks used in the field of deep learning and (Sze, Chen, Yang, & Emer, 2017) have detailed the hardware history and advances in the field.
There has been extended and comprehensive research in 2D vision tasks with some reaching maturity stage and producing state-of-the-art results (Voulodimos, Doulamis, Doulamis, & Protopapadakis, 2018). For a comprehensive review on deep learning, (Pouyanfar, et al., 2018) is a good place to look.

The success of deep learning in many 2D vision tasks led to attempts of using it in the 3D domain. But extending these models to 3D has proven to be difficult because of its different challenging properties. Such difficulties are complex geometric nature of its objects, different representations as seen in the last section results in large structural variations. The grid-like structure of 2D data makes it possible to use Deep Neural Networks (DNN) architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) because theywork best on structured data but some representations of 3D data e.g. point cloud and meshes are not structured and as such the use of such networks need a lot of finetuning. These difficulties consequently make conventional vision tasks like object detection and segmentation challenging.

As seen in Section III, 3D data has many representations that serve as input format for learning algorithms/architectures. The next section looks at the different categories of learning algorithms based on the type of input they consume.

a. Multi-view input: the use of multi-view data is one of the earliest and easiest and ways of extending 2D DL models to3D. It allows for 3D learning and understanding but from 2D point of view and at the same time taking into consideration the 3D geometry of an object. One of the standout works is Multi-View CNN (MVCNN) (Su, Maji, Kalogerakis, & Learned-Miller, 2015) which surpassed the performance of image classifiers in 2D. The architecture learns feature descriptors by supplying an ImageNet pre-trained VGG network(Simonyan & Zisserman, 2015) with images and feeding these features to CNNs for additional feature learning. 3D multi-view DL models can be used in extending already existing 2D DL paradigms without much finetuning. Other multi-view architectures are: Deep Belief Networks (DBN) for extracting features from depth images(Leng, Zhang, Yao, &Xiong, 2014)and CNNs for object recognition (Johns, Leutenegger, & Davison, 2016). A comparison between volumetric and multi-view CNNs was done by (Qi, et al., 2016)and they were able to outperform the state-of-the-art in 3D object classification in both.

   Multi-view image learning has many drawbacks and headmost is "pseudo-learning" because perception is done in 2D perspective which does not accurately depict the

innate 3D structure. Another disadvantage is the determination of the requisite number and methods of acquisition of the views to be used. These problems were motivation for further research on how to learn directly from 3D data.

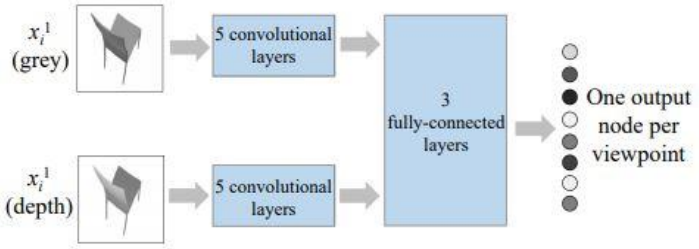

Figure 3: CNN-2 architecture for next-best view classification (Johns, Leutenegger, & Davison, 2016)

b. Volumetric/Voxel grid input: Even though multi-view architectures were less computationally extensive than volumetric architectures, the latter are the closest to representing 3D data. Its grid-like structure also makes it easier to extend 2D DNNs like CNNs. The earliest work to use voxel input is ShapeNet(Wu, et al., 2015) and comprises of one input layer, three convolution layers and one output layer. It was used on three tasks (3D shape classification, next-based view prediction and view-based recognition). Its main drawback was computational intractability. (Maturana & Scherer, 2015) proposed 3D convolutions in VoxNet with 32*32*32 voxel grid as input and it outperformed ShapeNet in classification when tested on ModelNet10, ModelNet40 and NYUv2 datasets. Truncated Signed Distance Function (TSDF) was used in creating voxels from scanned RGB-D scenes and fed into Deep Sliding Shapes model (Song & Xiao, 2016) to perform object recognition and classification on Model Net dataset. Results from vision tasks using volumetric input were better than those obtained from multi-view architectures but their major drawback is they are computationally expensive due to convolution on voxel grids which is much higher than in 2D images and an increased number of parameters. This was the motivation for LightNet (Zhi, Liu, Li, & Guo, 2017) and OctNet (Riegler, Ulusoy, & Geiger, 2017). The former reduces computational complexity by learning many features using multitasking and the use of batch normalization while the latter takes advantage of the sparsity of its input format to hierarchically divide the space with the help of unbalanced octrees which subsequently enables the model to allocate memory space and computational power to dense regions.

   Despite the attempts by (Riegler, Ulusoy, & Geiger, 2017) and (Zhi, Liu, Li, & Guo, 2017), volumetric architectures have less resolution than point clouds because they pack together closely-linked miniature features into one voxel. Point cloud architectures also consume less memory space than their volumetric counterparts because they directlyconsumethe data without transforming it to another format.

c. Point cloud input: The foremost work to directly consume point cloud is Point Net (Qi, Su, Mo, & Guibas, 2017) which uses the concept of Transformer Networks (TN) as used in (Jaderberg, Simonyan, Zisserman, & Kavukcuoglu, 2015) to curb the problem of transformation (translation and rotation) invariance inherent in point clouds. The output from the TNs is fed to an RNN module that sequentially learns features from the point

cloud while been invariant to the points. A permutation-invariant max pooling module aggregates all the learnt features. PointNet's performance in classification and segmentation fared well with state-of-the-art results. The TNs ensures PointNet's robustness to perturbation. The max pooling operation misses out on some features (very small) due tosummarization through aggregation and as such PointNet++ (Qi, Yi, Su, & Guibas, 2017) was proposed to take care of this problem through hierarchical feature learning. Another approach is the use of kd-tree structure for the input point cloud thereby getting rid of computational intractability of convolutions in images or other structured 3D data formats like voxel grids. Kd-network's (Klokov & Lempitsky, 2017) performance when tested on shape classification, shape retrieval and shape part segmentation was very impressive compared to state-of-the art. There have been attempts to address point cloud vision tasks using unsupervised learning, one such attempt is proposed in FoldingNet (Yang, Feng, Shen, & Tian, 2017).

Point cloud learning is an active research area with novel ideas emerging such as 2D-3D joint learning in SPLATNet (Su, et al., 2018) which achieved state-of-the-art performance in segmentation, RGB-D images combined with point clouds for object detection as in Frustrum Point Nets (Qi, Liu, Wu, Su, & Guibas, 2018), RNNs in segmentation in RSNet (Huang, Wang, & Neumann, 2018).

The inherent non-Euclidean structure of point clouds makes modeling the local dependencies of points challenging and is still the main problem that researchers are working on.

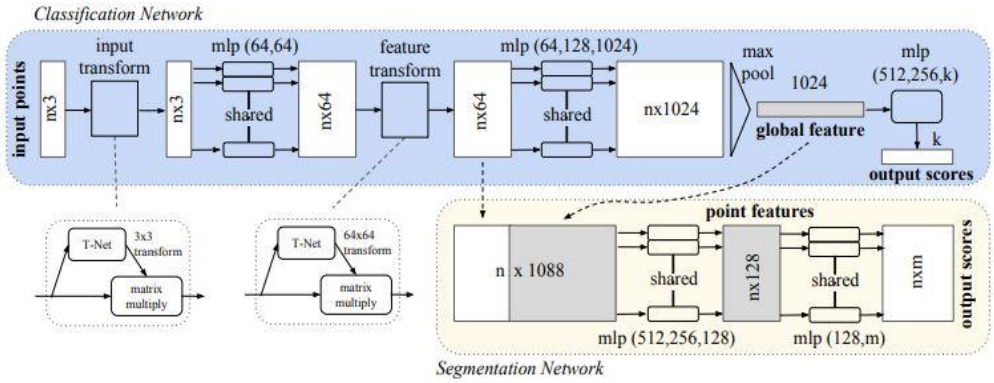

Figure 4: PointNet architecture (Qi, Su, Mo, &Guibas, 2017)

TRENDS AND FUTURE RESEARCH DIRECTIONS

There seems to be a shift from basic 3D processing for feature extraction, classification and segmentation like (Qi, Su, Mo, & Guibas, 2017) and (Qi, Yi, Su, & Guibas, 2017) to more complex tasks such as object detection like (Qi, Liu, Wu, Su, & Guibas, 2018), recognition and tracking. These tasks move the field closer to accomplishing vision perception.

Another promising branch is the use of Geometric Deep Learning (GDL) (Bronstein, Bruna, LeCun, Arthur, &Vandergheynst, 2017) to process graphs and meshes. Graphs and meshes can represent the innate structure of 3D data and both having non-Euclidean structure makes the

use of DNN on them a promising avenue to pursue. Works such as (Wang, et al., 2018) represented point cloud as a directed graph so as to capture the geometric dependencies between points. Another graph-based point cloud processing framework is (Landrieu & Simonovsky, 2018)'s use of super point graph which are geometric homogeneous partitions of the scanned point cloud scene. The network achieved a new state-of-the-art result in segmentation of LIDAR scans. The polygonal faces that makeup meshes can also be represented as graphs with the vertices equivalent to the graph's nodes and the connection between vertices representing edges. This property made it possible to extend graph-tailored CNNs and other DNNs to meshes. An example is the use of ReLu-NN in (Zhang, Li, Li, & Liu, 2018) for point cloud classification and reconstruction.

CONCLUSION

The increasing availability and affordability of 3D have given rise to a growing number of datasets which enables better performance of DNNs and other learning frameworks. Vision tasks like object classification, semantic segmentation, object detection and object tracking continue to produce promising results which proves that 3D data offer more accurate and discernable depiction of the world and relentless research in the direction is worthwhile. Advancements in this field will have a tremendous positive impact in autonomous driving, virtual reality, augmented reality, medical imaging, etc.

This paper has reviewed the methods of sensing in 3D and the different representations produced by the scanning devices. It also contains recent research on the different DL architectures employed in the field based on input format.

REFERENCES

Ahmad, E., Saint, A., Shabayek, A. E., Cherenkove, K., Das, R., Gusev, G.,...Ottersten, B. (2018, August 4). arXiv.org. Retrieved from arXiv.1808.01462:https://arxiv.org/abs/1808.01462

Aijazi, A. K., Checchin, P., &Trassoudaine, L. (2013). Segmentation Based Classification of 3D Urban Point Clouds: A Super-Voxel Based Approach with Evaluation. *MDPI Remote Sensing*, 1624-1650.

Blais, F. (2004). Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 231-243.

Bronstein, M. M., Bruna, J., LeCun, Y., Arthur, S., &Vandergheynst, P. (2017). Geometric Deep Learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 18-42.

Daneshmand, M., Helmi, A., Avots, E., Noroozi, F., Alisinanoglu, F., Arslan, H. S., . . . Anbarjafarim, G. (2018, January 24). *3D Scanning: A Comprehensive Survey.* Retrieved from arXiv.org: arXiv:1801.08863

Furukawa, Y., & Hernandez, C. (2015). Multi-View Stereo: A Tutorial. *Foundations and Trends in Computer Graphics and Vision*, 1-148.

Huang, Q., Wang, W., & Neumann, U. (2018). Recurrent Slice Networks for 3D Segmentation of Point Clouds. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2626-2635). Utah: IEEE.

Ioannidou, A., Chatzilari, E., Nikolopoulos, S., &Kompatsiaris, I. (2017). Deep Learning Advances in Computer Vision with 3D Data: A Survey. ACM Computing Surveys (CSUR)

Jaderberg, M., Simonyan, K., Zisserman, A., &Kavukcuoglu, K. (2015). Spatial Transformer Networks. *Neural Information Processing Systems (NIPS)*. Montreal: NIPS.

Johns, E., Leutenegger, S., & Davison, A. J. (2016). Pairwise Decomposition of Image Sequences for Active Multi-view Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3813-3822). Las Vegas: IEEE.

Klokov, R., &Lempitsky, V. (2017). Escape from Cells: Deep Kd-Networks for the Recognition of 3D Point Cloud Models. *International Conference on Computer Vision*, (pp. 863-872). Venice.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural networks. *International Conference on Neural Information Processing Systems (NIPS).I*, pp. 1097-1105. Nevada: Curran Associates Inc.

Landrieu, L., &Simonovsky, M. (2018). Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4558-4567). Utah: IEEE.

Leng, B., Zhang, X., Yao, M., &Xiong, Z. (2014). 3D Object Classification Using Deep Belief Networks. *International Conference on* (pp. 128-139). Dublin: Springer.

Mahony, N., Campbell, S., Krpalkova, L., & Riordan, D. (2018). Computer Vision for 3D Perception: A Review. Intelligent Systems Conference. London

Microsoft. (2018, November 17). Kinect - Windows app development. Retrieved from Kinect for Windows web site: https://developer.microsoft.com/en-us/windows/kinect

Maturana, D., & Scherer, S. (2015). VoxNet: A 3D Convolutional Neural Network for real-time object recognition. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 922-928). Hamburg: IEEE.

Nguyen, A., & Le, B. (2013). 3D Point Cloud Segmentation: A Survey. *IEEE Conference on Robotics, Automation and Mechatronics (RAM)*. Manila: IEEE.

Parvat, A., Chavan, J., Kadam, S., Dev, S., & Pathak, V. (2017). A Survey of Deep Learning Frameworks. *International Conference on Inventive Systems and Control (ICISC)*. Coimbatore: IEEE.

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., . . . Iyengar, S. S. (2018). A Survey on Deep Learning: Algorithms, Techniques, and Applications. *ACM Computing Surveys (CSUR)*, 92-118.

Qi, C. R., Liu, W., Wu, C., Su, H., &Guibas, L. J. (2018). Frustum PointNets for 3D Object Detection from RGB-D Data. *arXiv:1711.08488v2*. arXiv.org.

Qi, C. R., Su, H., Mo, K., &Guibas, L. J. (2017). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 652-660). Hawaii: IEEE.

Qi, C. R., Su, H., Niebner, M., Dai, A., Yan, M., &Guibas, L. J. (2016). Volumetric and Multi-View CNNs for Object Classification on 3D Data. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5648-5656). Las Vegas: IEEE.

Qi, C. R., Yi, L., Su, H., &Guibas, L. J. (2017). PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *Neural information Processing Systems (NIPS)*.擦力佛日: 泥菩萨.

Riegler, G., Ulusoy, A. O., & Geiger, A. (2017). OctNet: Learning Deep 3D Representations at High Resolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Hawaii: IEEE.

Sansoni, G., Trebeschi, M., &Docchio, F. (2009). State-of-The-Art and Applications of 3D Imaging Sensors in Industry, Cultural Heritage, Medicine, and Criminal Investigation. *Sensors*, 568-601.

Schoning, J., &Heidemann, G. (2016). Taxonomy of 3D sensors - A Survey of State-of-the-Art Consumer 3D-Reconstruction Sensors and Their Field of Applications. *Conference: Conference: International Conference on Computer Vision Theory and Applications (VISAPP)* (pp. 192-197). Rome: SCITEPRESS .

Simonyan, K., & Zisserman, A. (2015, April 10). arXiv.org. Retrieved from arXiv:1409.1556: https://arxiv.org/abs/1409.1556

Song, S., & Xiao, J. (2016). Deep Sliding Shapes for Amodal 3D Object Detection in RGB-D Images. *IEEE Conference on Computer Vision and Pattern (CVPR)*. Las Vegas: IEEE.

Su, H., Jampani, V., Sun, D., Maji, S., Kalogerakis, E., Yang, M.-H., & Kautz, J. (2018). SPLATNet: Sparse Lattice Networks for Point Cloud Processing. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Utah: IEEE.

Su, H., Maji, S., Kalogerakis, E. K., & Learned-Miller, E. (2015). Multi-view Convolutional Neural Networks for 3D Shape Recognition. *IEEE International Conference on Computer Vision (ICCV)* (pp. 954-953). Chile: IEEE.

Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proceedings of the IEEE*, 2295-2329.

Voulodimos, A., Doulamis, N., Doulamis, A., &Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, 1-14.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (2018, January 24). Dynamic Graph CNN for Learning on Point Clouds. *arXiv:1801.07829v1*. arXiv.org.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., & Xiao, J. (2015). 3D ShapeNet: A Deep Representation for Volumetric Shapes. *IEEE Conference on Computer Vision and Pattern recognition (CVPR)* (pp. 1912-1920). Boston: IEEE.

Yang, Y., Feng, C., Shen, Y., & Tian, D. (2017). FoldingNet: Interpretable Unsupervised Learning on 3D Point Clouds. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 206-215). Utah: IEEE.

Zhang, L., Dong, H., & El Saddik, A. (2015). From 3D Sensing to Printing: A Survey. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, Article 27, 23 pages.

Zhang, L., Li, Z., Li, A., & Liu, F. (2018). Large-scale urban point cloud labeling and reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 86-100.

Zhi, S., Liu, Y., Li, X., & Guo, Y. (2017). Toward real-time 3D object recognition: A lightweight volumetric CNN framework using multitask learning. *Computers and Graphics*, 199-207.

Zhu, L., Shen, S., Gao, X., & Hu, Z. (2018). Large Scale Urban Scene Modeling from MVS Meshes. *The European Conference on Computer Vision (ECCV)* (pp. 614-629). Munich: Springer.