Assessment of Methods for Industrial Indoor Object Recognition

Borhan SALEM[a], Josip STJEPANDIĆ[a,1] and Sebastian STOBRAWA[b]
aPROSTEP AG, Germany
bLeibniz University Hannover, Institute of Production Engineering and Machine Tools, Germany

Abstract. One of the most important means for digital transformation is the Digital Twin which is the most exact representation of the manufacturing process in a virtual model. The Digital Twin provides many advantages for the optimization of production processes, but it is currently still used very rarely. This is primarily due to the fact that companies are reluctant to incur generation costs and do not have the required IT know-how. One approach to overcome these obstacles is to simplify the generation of the Digital Twin by using object recognition. This can be used to automatically scan a production system and transfer it to a model. However, suitable methods of object recognition are needed to achieve added value. Suitable acquisition methods are also necessary to compensate the impact of darkness, dirt and occlusion. This paper provides a comprehensive overview of the recent advances in 3D object recognition of indoor objects using Convolutional Neural Networks (CNN). The comparison of main recognition methods based on methods of geometric shape descriptor and supervised learning and their strenghts and weakness is also included. The focus lies on the specific requirements and constrains in an industrial environment like tight assembly, light, dirt, occlusion, incomplete data sets.

Keywords. Object Recognition, Indoor Object Acquisition, Convolutional Neural Network, Simulation

Introduction

The digital transformation supports almost every aspect of our daily work in business. Most people have constant access to new technologies and use services accessible online. For people in the working age it is getting challenging to embrace work and live with old technology [1]. The important fact frequently overlooked is however that these services have only recently become available and are intended to improve at a rapid pace. The market share of products and services which are not yet affected by digital transformation is dramatically decreasing [2]. One of the most frequent questions in the field of digital transformation is how to handle the physical objects, the existing "hardware", which doesn't exist as an appropriate digital model? Such objects include complex long-living spatial objects like machines and industrial equipment in particular which are produced once and used in operation and maintenance for long time [3]. In the era of the Digital Twin, which simultaneously comprises the description of the real and virtual product, this requirement is of a particular importance [2].

---

[1] Corresponding Author, Email: josip.stjepandic@opendesc.com.

In the past few years, tremendous advancements have been achieved in 3D distance measurement equipment and low-cost 3D acquisition devices (e. g., Microsoft Hololens), which made 3D data easily accessible [4]. Significant research and development have been conducted in contactless 3D measurement, 3D object reconstruction, 3D object localization and 3D modeling [5]. Object recognition is the main task in 3D computer vision with various applications in medicine, robotics, augmented reality, and, finally, autonomous driving [6]. Deep learning (DL), mostly realized by the implementation of Convolutional Neural Networks (CNN), is the most important means for object recognition. Numerous deep learning (DL)-based 3D object recognition methods have been developed recently to achieve significantly better performance than the traditional handcrafted feature-based methods [7]. In an industrial environment, object recognition goes beyond the simple recognition of an obstacle in the space: each object must be put in the exact position and get the right orientation in space and identification attributes. Once the Digital Twin comprises the frequent change of the geometrical shape of an object, high semantical and geometrical accuracy in the process of the data acquisition is required. Subsequently, the high-quality CAD data of all geometrical objects in all stages of planning process are the pre-requisite for seamless downstream processes [8].
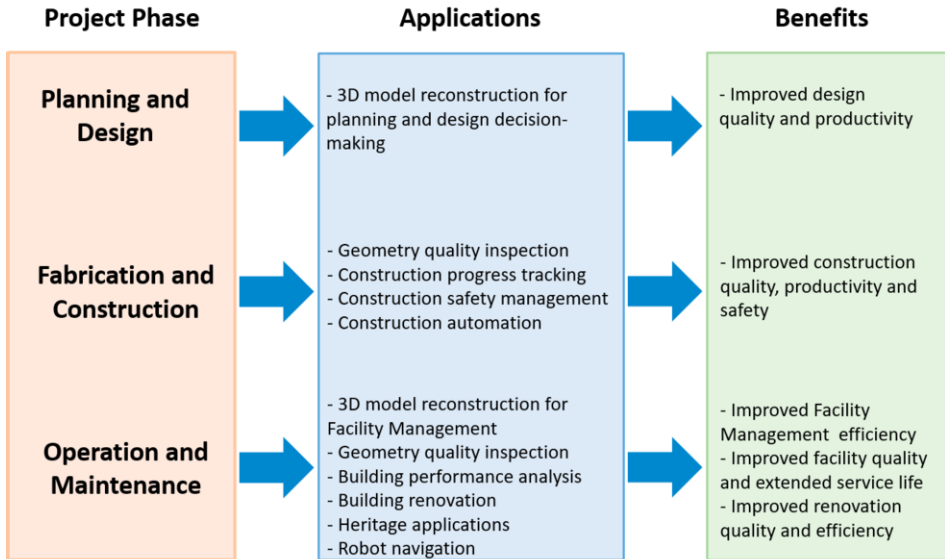
An object scanned by a laser is depicted as a point cloud. This dense point cloud consists of millions of points which contain information about space, color and intensity. Compared to photogrammetry, laser scanning is less sensitive to light conditions and produces a more accurate and detailed point cloud. However, the disadvantages are the high costs of the equipment and the time consuming data collection. Since production systems usually have occlusions, several scans from different positions are necessary. This is the only way to capture objects sufficiently and correctly [9].

Instead of identifying the detail feature of each object, the identification of CAD models from a reference library and the transfer of geometry and other object data (e. g. machine types) as modular objects directly from the library [10][11] to the design context can significantly reduce the scan times for a first rough "prescan" of the production. At the same time, database reconciliation enables the use of simpler and cheaper scanning methods [12][13]. The definition of suitable interfaces enables the transfer of information into a program for simulation of production systems and a precisely fitting Digital Twin of the manufacturing can be generated - almost without manual interventions [14][15]. The summary of applications and benefits of point cloud data in different project phases is given in Figure 1 [6].

Three contributions are made in this paper. First, an overview of the requirements for object recognition in industrial environments is given. The test basis is then presented and various criteria defined. Finally, single methods for object recognition are described exemplarily for the industrial environment. A large number of different methods for object recognition are currently being published in the scientific community [16]. There are many different approaches, like using voxels, image projections, raw point clouds and graphs. In almost all cases, CNN were trained with this data. It appears that Graph CNN have the potential to be a superior method. This is because they contain more information than point clouds or image projections at lower cost than voxels [17]. The applicability of different representations to unforeseen 3D point cloud data sets will be discussed and an approach for the implementation will also be presented.

The remainder of this paper is structured as follows: in Section 1, the background on reconstruction of semantic model is given, followed by the description of the test base in Section 2. In Section 3, the requirements for data acquision accuracy are highlighted.

Section 4 gives the comparison of methods with the related discussion in Section 5. Finally, Section 6 summarizes the conclusions and outlook.

Project Phase	Applications	Benefits
Planning and Design	- 3D model reconstruction for planning and design decision-making	- Improved design quality and productivity
Fabrication and Construction	- Geometry quality inspection - Construction progress tracking - Construction safety management - Construction automation	- Improved construction quality, productivity and safety
Operation and Maintenance	- 3D model reconstruction for Facility Management - Geometry quality inspection - Building performance analysis - Building renovation - Heritage applications - Robot navigation	- Improved Facility Management efficiency - Improved facility quality and extended service life - Improved renovation quality and efficiency

Figure 1. Usage of cloud point data in different project phases [6].
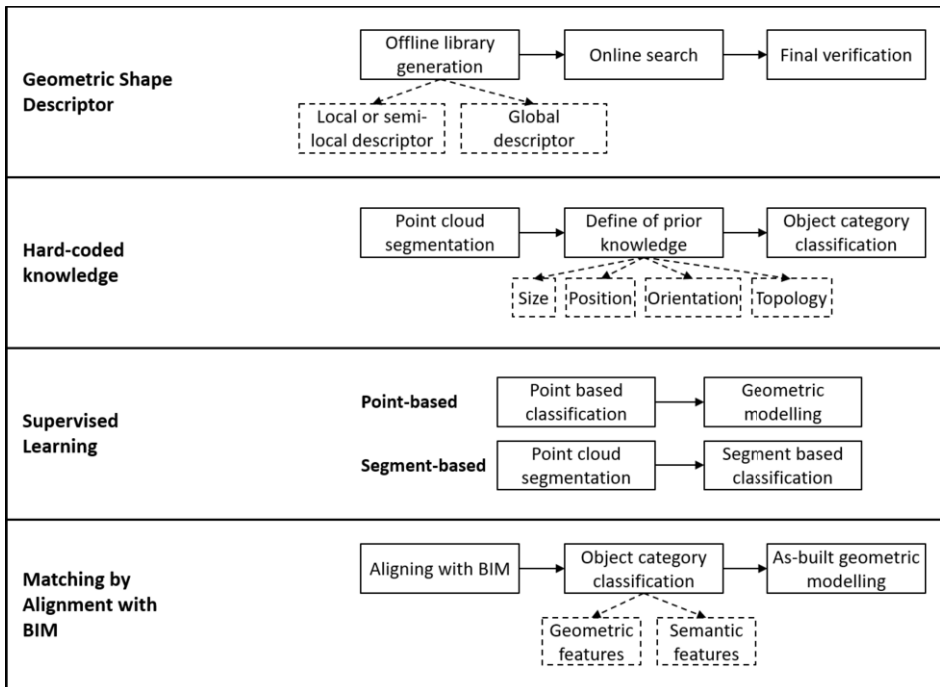
1. Reconstruction of semantic model

Digital Twin has few pre-requisites. For our research, the most important pre-requisite is the availability of high-quality models of all involved objects. In the case of 3D point cloud data in a factory, the two major applications are 3D model reconstruction and geometry quality inspection [16]. We will concentrate on the reconstruction because quality reconstruction looks less challenging in terms of recognition.

The two categories of 3D model reconstruction are distinguished by reconstructing either geometric or semantic information. In the first case, the model only contains the geometric information. The second category additionally contains object-based semantic information. Hence, the process involves the recognition of objects from point cloud data and the enrichment of semantic information [6]. In addition, the aim of object recognition here is to identify different object categories from the point cloud, such as windows or walls. For this purpose, different methods for object recognition were developed and presented in the literature. Examples are supervised learning and hard-coded knowledge.

The strategies and algorithms for semantic model reconstruction vary according to different problems. Based on the literature review [6], the four most popular approaches to semantic model reconstruction from point cloud data include: (1) geometric shape descriptor-based approach, (2) hard-coded knowledge-based approach, (3) supervised learning-based approach, and (4) matching by alignment with BIM based approach. The procedures of the four approaches are illustrated in Figure 2.

The geometric shape descriptor-based approach is suitable for cases where the objects to be detected possess a CAD/BIM model [18]. Different objects are shaped specifically, so a geometric shape descriptor that depicts the shape of an object is used to identify an object instance or object class [19]. As shown in Figure 2, the shape

descriptor-based method proceeds in three steps: (1) offline library generation, (2) online search and (3) final verification. The geometric shape descriptor-based approach is well suited for the search and recognition of complex geometric features with unknown positions from point cloud data [20]. However, the approach is only applicable if the CAD/BIM model of an examined object is available.

Figure 2. Four different approaches to semantic model reconstruction [6].

The hard-coded knowledge-based method is an effective algorithm for simple geometric components with different geometric features. However, the algorithm has narrow constraints due to hard-coded knowledge, so the algorithm only works for simple geometries. For example, the algorithm runs stable when detecting walls or roofs.

The algorithm uses machine learning to train classifiers. With this supervised learning-based approach point clouds are mapped to predefined object categories. There are two strategies for supervised learning-based object recognition, namely point-based classification and segment-based classification [6].

The supervised learning-based approach generally has wide applicability. With this method, it is possible to recognize objects with complex structures and properties, such as construction equipment and scaffold. The disadvantage of the method, however, is that a large amount of labelled data is needed to train the classifier. This leads to a large effort for the setup of the method, which is also time-consuming [6].

2. Test base

A critical aspect in the procedure for object recognition is the test base. Since this aspect usually appears in every object recognition project, a large number of 3D data sets have

been published in recent years. These data sets can be used to train the classifier. There are many sets available for indoor and outdoor applications [21]. However, these data sets are general and the test base will never fully fit the specific problem. Accordingly, appropriate training with suitable training data is essential. This is the only way to achieve high accuracy and low overfitting at the same time [22]. Table 1 shows a list of different 3D data sets. These focus on the object recognition of single objects. Here the data set ModelNet [23] is one of the most used for the research of 3D object recognition. It contains two splits, ModelNet10 and ModelNet40, which are used to classify household items such as chairs and desks [17].

Table 1. Popular 3D data sets that focus on studying individual objects along with their number of samples and classes.

Data set	#Samples	Classes
ModelNet10	4,899	10
ModelNet40	12,311	40
ShapeNet Core	51,300	55
Sydney Urban Objects	588	14

ShapeNet is a large 3D database developed for 3D segmentation and reconstruction. It contains a large number of labelled records of objects. Therefore, its use supports the task and it can be used to train the classifier [17]. The fourth data set, Sydney Urban Objects, is a small point cloud data set. The point clouds were measured by a LIDAR sensor that examines an urban environment. In contrast to the other data sets mentioned above, the data is taken from reality and is thus filled with noise. This, however, comes closer to the real application case and is therefore preferred [17].

There are many studies in the literature that use ModelNet as a test base for 3D classification. Here, the 3D meshes are often converted into voxels because they can be convolved like any other tensor. Estimating the voxel orientation in addition to the classification performed can improve the overall results [17].

3. Requirements for data acquision accuracy

Prior to each process and method development, a detailed analysis of the requirements for accuracy is necessary. Each factory and each building have their own history and life cycle. Most commonly, the status as-built is poorly documented and must be thorougly captured with sufficient accuracy in case of change, reorganization and retrofit [5].

Although various parameters have been proposed, the two most common parameters for point cloud data quality are accuracy and spatial resolution.The parameter accuracy measures the ranging precision of each single point. This is mainly determined by the hardware performance of the equipment as well as the data acquisition parameters [6].

The other parameter, spatial resolution, measures the spatial density of point cloud data. It is mainly determined by the data acquisition parameters. In some cases, coverage is mentioned as another parameter of point cloud data quality [6]. The determination of parameters for point cloud acquisition is necessary to fulfill the required data quality and to carry out the acquisition in a time-effective manner [6].

In the current state, the point cloud is created with a high manual effort. The expert knowledge and experience of the engineers is also needed here. On the one hand, the

collection of insufficient data makes a model useless for the intended application. On the other hand, the collection of too much data requires more time and effort and leads to redundant data. Accordingly, it is important to find a suitable degree and to determine the required quality of the point cloud for each specific application. With this, the acquisition of point cloud data can be facilitated [6]. By using a modern terrestrial scanner, a scan with subsequent automatic registration takes approximately twelve minutes.

4. Comparison of methods

For the recognition of a dedicated object in a section of a factory, deep learning (DL)-based methods are used. DL techniques have emerged as powerful methods for learning features automatically from data [24]. In particular, these techniques have provided significant improvement for object detection, a problem which has attracted enormous attention in the last few years, even though it has been studied for decades before [25].

Basically, two different methods are used for objects recognition in point cloud data gathered with a scanner: point-based and voxel-based. However, there are certain difficulties in using DL to process 3D point clouds [21] which must be taken into account:

− The point cloud is not in a regular format. Unlike an image, a point cloud is a set of points distributed in space. Hence, it is non-grid data. However, typical convolutional architectures can merely deal with highly regular data formats, and it is, therefore, hard to use Convolutional Neural Networks to process raw point clouds [22].
− The point cloud is unordered. There are various orderings of points, indicating that there are many matrix representations of a particular point cloud. Therefore, adapting the changes of how the waypoints are arranged is a problem in point cloud processing [22].
− Three-dimensional point clouds normally contain only the spatial coordinates of points, lacking rich textures, colors, and other information.

To resolve these disadvantages, point clouds can be translated into voxels for further processing. However, this process chain pre-requisites several formal rules and quality of point cloud. There are two downsides to voxels. First, 3D meshes and point clouds only encode the surface, so most of the voxels in a sample are zero valued. Second, the dimensionality of voxels increases cubically, which is why most applications keep the dimensionality small (e. g. $32 \times 32 \times 32$), which eliminates most fine details [17]. For this reason, the distinction of two almost indentical objects with small differences could be a serious issue.

In terms of implementation, some practical questions arise. While multiple dozens of similar approaches with similar accuracy exist and are being developed further, a practical implementation needs a selection based on few most important criteria. These are (not exhaustive): (a) high accuracy, (b) robustness, (c) low parameter number, (d) the stable, open-access framework proven by a large community at GitHub, (e) real-time execution (if possible), (f) low resource demand (standard PC). Obviously, there is a conflict of goals and the selection must be done using the best compromise.

The most interesting approaches and their main characteristics are shown in Table 2. Geometric shape descriptors (Principal Axle Descriptor [26] and Variational Autoencoder [27]) build a vector and search for similar objects in a library. G3DNet

[17] and PointNet [28], as point-based methods, have a DL architecture that is used in 3D object classification and segmentation. It takes point clouds directly as inputs and outputs. In contrast, a 3D detection framework (VoxNet, VoxelNet [29], LightNet [30]) transforms the point cloud into voxels that each contain a small amount of points. It produces bounding boxes based on the features of the voxels. Finally, a combined approach, called FusionNet [31], uses both volumetric and pixel information to recognize 3D objects. A similar approach with face normals is used by NormalNet [32]. For the sake of completeness and comparison, 3DShape Nets and ObjectNet3D are also added [12].

**Table 2.** Main characteristics of different approaches for object recognition.

Approach	Input	Frame-work	Open source	Test Data	No. Params	Augmen-tation	Accur. %	Publi-shed in
VoxNet	Voxel	Theano/ Lasagne	Y	MN 40	0.92 M	12 rot	83.00	2018
Principal Axle Descriptor	Voxel	Scikit-Learn	N	Own		N/A	93.18	2018
VoxelNet	Voxel	Tensor-flow	Y	KIT TI		12 rot	89.35	2018
G3DNet	Point Cloud	Tensor-flow	Y	MN 40		N/A	91.00	2018
PointNet	Point Cloud	Tensor-flow	Y	MN 10	80 M	N/A	77.60	2017
FusionNet	Voxel + Pixel	Caffe	N	MN 40	118 M	60 rot	90.80	2016
NormalNet	Normal vector + voxel	Tensor-flow	N	MN 40	6.5 M	20 rot	88.80	2019
Variational A-encoder	Voxel	Theano/ Lasagne	Y	MN 40	18 M	24 rot	88.98	2016
LightNet	Voxel		N	MN 40	0.3 M	12 rot	88.93	2018
3DShape Nets	Depth Map		N	MN 40	38 M	12 rot	77.30	2015
Object Net3D	Images		Y	Obj Net			98.90	2016

The rough overview in Table 2 shows disparate characteristics of pre-selected approaches. An accuracy in range of ninety percent can be expected with quite disparate approaches. A trend for further development is not apparent. The combination of two patrallel, different approaches in one application to prevent overfitting may be the better option.

At this time, the real-time processing and recognition is a desirable option, but the expense for training can be hardly pre-estimated.

5. Discussion

Deep learning with 3D data has not been researched nearly as much as it has been with images. In comparison to object recognition methods based on images or depth maps (e. g. Object Net3D [12)], which achiev an accuracy of almost hundred percent, 3D point cloud data based methods still require further development and provide higher potential for many applications (e. g. harsh environments). Main obstacles are both the recognition methods and the extremely time-consuming handling of the huge volume of data. Therefore, it is necessary to select several frameworks for 3D object recognition which promises the best results for investigated objects in the factory for different use cases. The future provision of a specific benchmark data set for "rough environments", which comprises typical industrial object like machines, elevators, structural elements, pipelines, etc., by a standardization body or industrial consortium would also facilitate such a research.
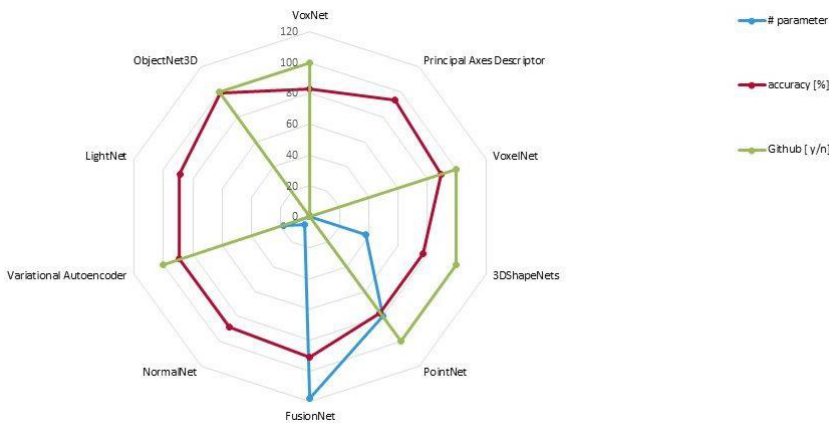
Figure 3. Comparison of main object recognition approaches for practical implementation.

Even the basic choice between the geometric shape descriptor and supervised learning does not give a clear result (Table 2 and Figure 3) nor a trend of future progress. None of the methods provide a sufficient solution for specific practical issues: different working stati (e. g. robots), occlusion, incomplete data sets, indesired reflections in scans on glass planes, etc..

Like most other object recognition architectures, voxel-based approaches are not inherently rotation-invariant. For this reason, a data augmentation technique needs to be implemented. During training, each model is rotated twelve times and trained on all copies. At test time, the output of the final fully-connected layer is pooled across several rotations of the input. In this way, voxel-based approach learns rotation invariance by sharing the same learned convolutional kernel weights across different rotations of the input voxel grid [32][33].

Voxel grids still have a number of drawbacks. They lose resolution compared to point clouds, as several distinct points representing intricate structures will be binned into one voxel if they are close together. Voxel grids can yield unnecessarily high memory usage compared to point clouds in sparse environments, since they actively use memory to represent free and unknown space whereas point clouds contain only known points [33]. Finally, if an object is recognized, it must be put in the right position and proper orientation in the space.

6. Conclusions and outlook

The intention of this paper to close the gap between the real and virtual factory in the modern manufacturing industry by the generation of a Digital Twin of a built environment is presented. Different applications in other areas, such as autonomous driving, mean that the field of 3D point cloud sensors is experiencing great technical progress. In contrast to image, video, and audio data, which occur on regular gridded structures, 3D point cloud data requires special purpose convolution and pooling operations. For this purpose, dozens of methods based on geometric shape descriptor as well as supervized learning are available. Methods for automatic object recognition based on geometric shape descriptor and supervised learning have reached a level of maturity that makes them suitable for industrial practice. Based on ModelNet40 data set, accuracy of more than ninety percent can be expected and almost every month a new method published. Selection of a suitable method for practical implementation is difficult because essential parameters differ greatly for individual approaches.

The aim of this study is to develop an industrial approach for a commercial service to be implemented by www.OpenDESC.com to generate a Digital Twin based on fast scanning of a factory and subsequent object recognition.

Acknowledgement

References

[1] R.Y. Zhong and W. Ge, Internet of things enabled manufacturing: a review, *International Journal of Agile Systems and Management*, Vol. 11, 2018, No. 2, pp. 126-154.

[2] N. Wognum, C. Bil, F. Elgh, M. Peruzzini, J. Stjepandić and W.J.C. Verhagen, Transdisciplinary systems engineering: implications, challenges and research agenda, *International Journal of Agile Systems and Management*, Vol. 12, 2019, No. 1, pp. 58-89.

[3] J. Ríos, F.Mas Morate, M. Oliva and J.C. Hernández, Framework to support the aircraft digital counterpart concept with an industrial design view, *International Journal of Agile Systems and Management*, Vol. 9, 2016, No. 3, pp. 212–231.

[4] S. Bondar, B. Salem and J. Stjepandić, Indoor Object Reconstruction based on Acquisition by Low-cost Devices, *Advances in Transdisciplinary Engineering*, Vol. 7, 2018, pp. 113-122.

[5] Z. Ma, S Liu, A review of 3D reconstruction techniques in civil engineering and their applications, *Advanced Engineering Informatics*, Vol. 37, 2018, pp. 163–174.

[6] Z. Geng and B. Bidanda, Review of reverse engineering systems – current state of the art, *Virtual and Physical Prototyping,* Vol. 12, No. 2, pp. 161–172.

[7] S. Singaravel, J. Suykens, P. Geyer, Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction, *Advanced Engineering Informatics*, Vol. 38, 2018, pp. 81–90.

[8] S. Bondar, C. Ruppert and J. Stjepandić, Ensuring data quality beyond change management in virtual enterprise, *International Journal of Agile Systems and Management*, Vol. 7, 2014, Nos. 3/4, pp. 304–323.

[9] J. Chen, Y. Fang, Y. Cho and C. Kim, Principal Axes Descriptor for Automated Construction-Equipment Classification from Point Clouds, *ASCE's Journal of Computing in Civil Engineering,* 2016, 10.1061/(ASCE)CP.1943-5487.0000628.

[10] H. Fathi, F. Dai, M. Lourakis, Automated as-built 3D reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges, *Advanced Engineering Informatics*, 2015, Vol. 29(2), pp. 149-161.

[11] J. Stjepandić, E. Ostrosi, A.-J. Fougères, M. Kurth, Modularity and Supporting Tools and Methods, in: J. Stjepandić et al. (eds.): *Concurrent Engineering in the 21st Century: Foundations, Developments and Challenges*, Springer International Publishing Cham, 2015, pp. 389–420.

[12] V. Pătrăucean, I. Armeni, M. Nahangi, J. Yeung, I. Brilakis and C. Haas, State of research in automatic as-built modelling, *Advanced Engineering Informatics*, Vol. 29 (2) , 2015, pp. 162-171.

[13] A. Biahmou, C. Emmer, A. Pfouga and J. Stjepandić, Digital master as an enabler for industry 4.0, *Advances in Transdisciplinary Engineering*, Vol. 4, 2016, pp. 672-681.

[14] C. Emmer, A. Fröhlich and J. Stjepandic, Advanced engineering visualization with standardized 3D formats, *IFIP Advances in Information and Communication Technology*, Vol. 409, Springer, Berlin Heidelberg, 2013, pp. 584-595.

[15] N. Bicocchi, G. Cabri, F. Mandreoli and M. Mecella, Dynamic digital factories for agile supply chains: An architectural approach, *Journal of Industrial Information Integration*, 2019, in press.

[16] K. Wang, M.-K. Kim, Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018, *Advanced Engineering Informatics*, Vol. 39, 2019, pp. 306–319.

[17] M. Dominguez, R. Dhamdhere, A. Petkar, S. Jain, S. Sah, and R. Ptucha, General-Purpose Deep Point Cloud Feature Extractor, *2018 IEEE Winter Conference on Applications of Computer Vision*, pp. 1972-1981.

[18] Chen, J., Fang, Y., and Cho, Y., Performance Evaluation of 3D Descriptors for Object Recognition in Construction Applications, *Automation in Construction*, Vol. 86, Feb. 2018, pp. 44-52.

[19] X. Zhao and H.T. Ilieş, Learned 3D shape descriptors for classifying 3D point cloud models, *Computer-Aided Design and Applications*, Vol. 14, No. 4, pp. 507-515.

[20] X. Liu, Y. Lu, T. Wu and T. Yuan, An improved local descriptor based object recognition in cluttered 3D point clouds, *International Journal of Computers, Communications and Control*, Vol. 13, Issue 2, 2018, pp. 221-234.

[21] Le Zhang, Jian Sun, and Qiang Zheng, 3D Point Cloud Recognition Based on a Multi-View Convolutional Neural Network, *Sensors*, 2018, 18, 3681; doi:10.3390/s18113681.

[22] J. Dekhtiar, A. Durupt, M. Bricogne, B. Eynard, H. Rowson, D. Kiritsis, Deep learning for big data applications in CAD and PLM – Research review, opportunities and case study, *Computers in Industry*, Vol. 100, Sep 2018, pp. 227-243.

[23] A. Notchenko,Y. Kapushev and E. Burnaev, Large-Scale Shape Retrieval with Sparse 3D Convolutional Neural Networks, in W. van der Alst (ed.) *International Conference on Analysis of Images, Social Networks and Texts*, Springer, London, 2017, pp. 245-254.

[24] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu and M.S. Lew, Deep learning for visual understanding: A review, *Neurocomputing*, Vol. 187, 2016, pp. 27-48.

[25] G. Šagi, Z. Lulić and I. Mahalec, Reverse Engineering, in: J. Stjepandić et al. (eds.) *Concurrent engineering in the 21st century: Foundations, developments and challenges*, Springer International Switzerland, 2015, pp. 319-353.

[26] D. Maturana and S. Scherer, VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition, *IEEE International Conference on Intelligent Robots and Systems*, 2015-December,7353481, pp. 922-928.

[27] A. Brock, T. Lim, J.M. Ritchie and N. Weston, Generative and Discriminative Voxel Modeling with Convolutional Neural Networks, *3D Deep Learning Workshop at NIPS 2016*, 2016, http://3ddl.cs.princeton.edu/2016/papers/Brock_et_al.pdf, accessed June12 2019.

[28] C.R. Qi, H. Su, K. Mo and L.J. Guibas, PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* Honolulu, HI, USA, 21–26 July 2017.

[29] Y. Zhou, O. Tuzel, VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4490-4499.

[30] S. Zhi, Y. Liu, X. Li, and Y. Guo, LightNet: A Lightweight 3D Convolutional Neural Network for Real-Time 3D Object Recognition, *Eurographics Workshop on 3D Object Retrieval*, 2017, pp. 9-16.

[31] V. Hegde and R. Zadeh, FusionNet: 3D Object Classification Using Multiple Data Representations, *3D Deep Learning Workshop at NIPS 2016*, https://stanford.edu/~rezab/papers/fusionnet.pdf, June 12, 2019

[32] C. Wang, M. Cheng, F. Sohel , M. Bennamoun, J. Li, NormalNet: A voxel-based CNN for 3D object classification and retrieval, *Neurocomputing*, Vol. 323, 2019, pp. 139–147.

[33] V.V. Lehtola, H. Kaartinen, A. Nüchter, R. Kaijaluoto, A. Kukko, P. Litkey, E. Honkavaara, T. Rosnell, M.T. Vaaja, J.-P. Virtanen, M. Kurkela, A. El Issaoui, L. Zhu, A. Jaakkola and J. Hyyppä, Comparison of the Selected State-Of-The-Art 3D Indoor Scanning and Point Cloud Generation Methods, *Remote Sensing*, 2017, 9, 796; doi:10.3390/rs9080796.