Point Cloud Capturing and AI-based Classification for as-built BIM using Augmented Reality

Thomas Klauer and Bastian Plaß

i3mainz, Institute for Spatial Information and Surveying Technology,
Mainz University of Applied Sciences
(thomas.klauer, bastian.plass)@hs-mainz.de

Abstract. The benefits of using Building Information Modeling (BIM) have been proven in architecture, engineering and construction industry. However, implementing BIM in in facility management has not been achieved yet due to missing complete and accurate as-built BIM. Modeling comprehensive information for as-built documentation from 3D point cloud data is referred as Scan-to-BIM but lacks automation caused by unstructured data and high user input. We tackle the main issue of structuring the 3D point cloud data by using artificial intelligence while capture. With both, a highly reliable and low-cost technology we achieve less time-consuming point cloud capturing and segmentation contributing to a novel Scan-to-BIM approach with promising initial results.

Keywords: LiDAR; Point Cloud; Classification; Augmented Reality; Scan-to-BIM.

1 Introduction

Similar to many other sectors, digitalisation is advancing rapidly in the architecture, engineering and construction (AEC) industry. An important component here is the concept of the "digital twin" of a new building to be constructed or an existing building to be operated on. Building Information Modelling (BIM) has been established as the method for this, in which a 3D building model represents the core element. There are various ways to create such a model: in the case of new buildings to be planned, the model is generated "form the scratch" by the planners with specialised design software, while in the case of existing buildings it is necessary to capture the real building geometry with its relevant component information in reality. In addition to planning and construction, the management of existing buildings also benefits from digital BIM solutions, so that facility management will be able to exploit the advantages of BIM in the future with the ongoing development of efficient solutions for digitising existing buildings.

There are various approaches to 3D as-built modelling, such as deriving from 2D CAD plans (as-planned BIM) or capturing the up-to-date building representation by metrological methods (as-built BIM). In the latter, state-of-the-art 3D point cloud data are obtained from laser scanning or structure from motion (SfM) methods and serve both registered and manually pre-processed as the information basis for modelling a semantically enriched as-built BIM (Scan-to-BIM). The as-built modelling process lacks automation yet due to missing, sparse, outdated and complex information about the captured objects, relationships and attributes as well as customisable uses of the BIM [1,3-4]. In addition, the use of professional and therefore expensive metrology hardware has been required to perform such scans and specialised experts are needed to carry out both the scanning and modelling processes. This paper will show how these processes can be improved in terms of automation and simplification.

For an initial simplification of the scanning process, preliminary work [2] has shown that sufficient point cloud quality can be achieved with inexpensive consumer products such as the

Apple iPad Pro or the Intel RealSense L515 for as-built modelling of indoor scenes. This hardware can be handled by non-experts after a brief instruction, resulting in a significant cost reduction for capturing the 3D building geometry. In order to automate and thus simplify the modelling process, methods are needed that can provide semantically structured information of the scanned building geometry to identify relevant, constituent objects, such as building components, furnishings or building services elements.
One way to achieve this is to divide the raw and unstructured 3D point cloud into semantic regions by means of segmentation, in which objects are then automatically recognised [3] and finally geometrically approximated by standard geometries. Structuring the 3D data into semantic regions for further understanding thus represents the initial technical step in the BIM modelling process (c.f. Fig. 1). Artificial Intelligence (AI) methods such as machine learning (ML) can be used for segmentation and classification of 3D point cloud data showing various characteristics and dealing with different conditions usually. Published research [1,4-5] confirms the successful use of automated approaches for highly simplified building representations, but less so for the representation of complex reality.

This paper presents the development of an intelligent 3D data acquisition and processing method using LiDAR-based consumer hardware, designed for Apple's Pro Series such as the iPhone 12 Pro and the iPad Pro. The prototyped 3D data application called "Semantic Data Capture" provides a detailed and semantically structured point cloud using AI based on high-resolution depth data acquired by Apple's vision technology without prior calibration and less technical knowledge. This is done by capturing a depth map with the mobile device's built-in sensors. In combination with MLCore, extended by a third party ML model, captured geometries can be classified simultaneously into pre-defined building component categories. The prototype also uses Apple's augmented reality framework (ARKit), which allows users to visualise the results of the captured and classified data in the overlaid AR-image of the integrated RGB-camera synchronously. An application prototype has been developed that is usable also for non-experts, able (a) to generate detailed geometric representations of interior scenes, (b) to combine them with semantic attributes in real time and (c) to deliver a structured 3D point cloud for the further Scan-to-BIM process. As an application example, an interior analysis in the care sector was chosen here, which, for example, checks the suitability of living sites for certain diseases or care situations. A key feature of this novel application in academia is the simultaneous geometric capture and AI-based 3D data classification through a low-cost optical technology with direct visualisation for non-expert users, which has the potential to establish a new state-of-the-art Scan-to-BIM method.

2 Requirements for Scan-to-BIM and why it lacks automation

Referred as Scan-to-BIM, the automated reconstruction of existing buildings for BIM modelling is based on 3D point clouds, acquired by consolidated techniques such as laser scanning or SfM. Both techniques allow a rapid and up-to-date acquisition of as-built components with high spatial resolution but produce a huge amount of data as a consequence, that needs to be processed in a time-consuming and almost entirely manual process chain as presented in Fig. 1. The BIM model generation from acquired point cloud data can be roughly organised into the four main categories data preprocessing, segmentation, classification and BIM modelling.
Representing indoor scenes full covered by point clouds requires a variety of scan stations and viewpoints that need to be registered immediately after acquisition. The scope of the

preprocessed step is to remove outliers, smooth noise effects and for example downsampling or transforming the point cloud data into a usable format for subsequent processes [6]. Following that, the segmentation serves as the first technical step to transform the unstructured point cloud into several subsets according to the semantic property of points with respect to the scene characteristic. In line with the subsequent demand, the subsets can address rooms, unique constituent elements or specific regions of interest such as interiors or furnishing elements. Aside the segmentation, classification becomes relevant for mapping the segmented but disordered point cloud data by feature extraction and component identification into regular forms that are capable for the final BIM modelling.

Fig. 1. Scan-to-BIM stages according to [6]. The presented approach summarises the first three categories in order to achieve automation and simplification. Consequently, it reduces the manifold process chain for as-built BIM reconstruction.

Despite the increasing popularity of BIM for years and a quantity of technical solutions for automated Scan-to-BIM, the overall process remains costly and manually. Recent research efforts have been made in the development of automated point cloud segmentation and classification methods, mostly using ML approaches based on artificial neural networks and deep learning [7-12]. However, point cloud processing is still in its infancy because of several challenges described below.

As a consequence of evolving 3D data acquisition technologies, point clouds represent the state-of-the-art in surface reconstruction for engineering and 3D modeling tasks in the last decades [13]. Given a number of discrete points that sample a surface, point clouds are used prevalently but their processing still faces many challenges due to data imperfections and a variety of special data structure characteristics. The most challenging properties dealing with point clouds are the arbitrary sampling density that results in high redundancy on the one hand but still missing, sparse or obsolete data on the other hand due to occlusions, clutter and noise [1,4-5]. Apart from that, 3D point clouds have no inherent order such as pixel neighbours. That is why permutation invariance is another main property to take care of. Furthermore, point clouds are often dense with millions of single points that will provoke big data and large computational times raising cubicly on the number of point instances. All these properties affect the automatic processing of point clouds. Nevertheless, many efforts have been made to improve the quality of automated point cloud processing and to detach manual work. Even if the methods differ in approaches according to [14,15], they all used to apply after acquisition in postprocessing. Considering the large amount of data, postprocessing methods require the point cloud to be patched or significantly downsampled resulting in a loss of information and additional cost. While performing segmentation tasks on 2D image data using DL techniques represents state-of-the-art, a transfer to 3D point cloud data lacks due to above-mentioned reasons. Therefore, this paper proposes a new approach for point cloud segmentation in real-time during the acquisition process using deep learning techniques for both data paradigms, 3D mesh classification and reprojected 2D object classification result by a pre-trained model. Further details about our approach that is also using Augmented Reality (AR) for result visualisation are given in the following chapter.

3 Mobile AI-based Augmented Reality Approach

Mobile devices, such as tablets and especially smartphones with integrated LiDAR sensors, are in principle well suited for the 3D acquisition of buildings and building components, as they are widely used as personal devices and thus the users are usually familiar with their functionality and usability. Compared to professional acquisition hardware, they are also comparatively cheap to purchase and operate with and can be used also by non-experts. Just a few apps like[1,2,3,4] on the consumer market enable 3D scanning of indoor spaces. However, these apps are not able to carry out the necessary process steps such as segmentation and classification after the pure geometric acquisition. In order to avoid processing steps with a separate software after or detached from the capture, the main scope of the concept presented here was to integrate this directly into the mobile application, hereinafter app. The "Semantic Data Capture" app introduced here integrates the functionalities pointed out in Tab. 1.

Tab. 1. Functionality of the presented approach based on a mobile application.

Capturing the 3D geometries of inner building structures and interiors.
Segmentation, classification and recognition of components and furnishings.
Immediate visualisation of the capture and recognition with corresponding usability.
Saving of the results.
Exporting the results in relevant and open 3D formats.

3.1 Geometry Capturing

On a current mobile device from Apple, geometry capturing and also parts of object recognition are performed with the ARKit framework [16] using the built-in LiDAR sensor. In an AR session, ARKit stores all information that belongs to the captured environment. The core functionalities are tracking, i.e. the ability to follow objects relative to the position of the device, and scene understanding, i.e. the ability to collect information about the detected objects. In addition, ARKit offers simple integration into existing visualisation libraries (e.g. SceneKit and SpriteKit) as well as into individual visualisation solutions with Metal, Apple's computer graphics API. ARKit uses the built-in sensors of the mobile device such as the namely HD camera, 6D sensor for rotation, position and accelerometer as the basis for so-called Visual Inertial Odometry (VIO), for e.g. the determination of position and orientation supported by the RGB camera feed. The captured optical data is superimposed with the other device sensors (e.g. gyroscope and accelerometer) and then position and movement of the device in 3D space are calculated.
A method called raycasting is used to capture the spatial geometry, which allows 3D scenes to be displayed quickly. This is done by using a virtual ray that is projected from a point on the screen into the real world and allows the calculation of the intersection point with real objects (c.f. Fig. 2b). During the ongoing capture, ARKit then creates a virtual world of the captured geometries. To ensure that the computer-generated elements remain in their real positions, so-called virtual anchors are generated. The anchors used in the app presented here contain 3D geometry data that describe the objects from the environment in the form of nodes, polygons and

[1] 3d Scanner App: https://apps.apple.com/de/app/3d-scanner-app/id1419913995
[2] Canvas: Pocket 3D Room Scanner: https://apps.apple.com/us/app/canvas-pocket-3d-room-scanner/id1514382369
[3] Trnio 3D Scanner: https://apps.apple.com/de/app/trnio-3d-scanner/id683053382
[4] Capture: 3D Scan Anything: https://apps.apple.com/de/app/capture-3d-scan-anything/id1444183458

normals, which in turn form a polygonal mesh. In addition, semantic information can be predicted and assigned to each polygon from the mesh. Eight classes are currently supported with ARMeshClassification [17], namely ceiling, door, floor, seat, table, wall, window and *none,* when ARKit cannot predict the class of the polygon. Since these eight classes are not sufficient to recognise interiors with various other components and furnishings, an AI-based extension has been developed.

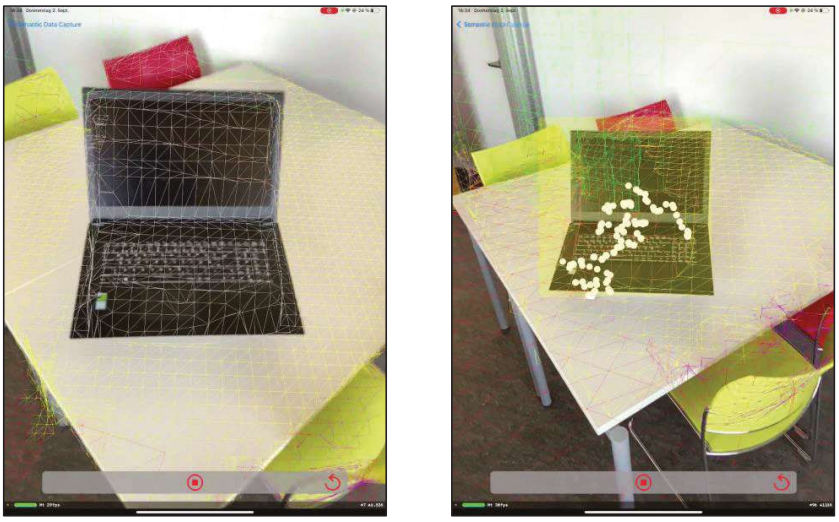
3.2 AI-based Mesh Classification and Object Detection

In order to be able to recognise objects in the virtual representation of a captured interior that are not part of the ARMeshClassification classes, i.e. they have been assigned to the class *none*, an AI-based extension was designed and prototyped. Various libraries or models are available here that can detect objects from (moving) images in real time. The YOLO (You Only Look Once) model was chosen in this work [18]. The approach of YOLO, in contrast to many other systems that work based on Convolutional Neutral Networks (CNN), is to perform object detection in a single pass – hence "you only look once". To make this possible, the CNN YOLO was trained with data from Microsoft's Common Objects in Context (COCO) database [19]. This trained network can now be applied to images or videos to perform multiple object detection in a fast manner. The model recognises features across the entire image and creates individual bounding boxes that assign a class to recognised objects according to the highest probability. Images are divided into a symmetric grid, where frames are suggested from each cell. Class probabilities are also calculated per cell, corresponding to the number of known classes in the training dataset. The class probabilities depend on the probability that an object is present in the cell.
The captured images are first preprocessed with Apple's vision framework. Afterwards a collection of the objects found is returned in the form of observations or an empty array if no objects were found by ARMeshClassification. An observation contains the class of the object and the normalised coordinates of the origin as well as the width and length of a frame within which the object should be located. This data is then passed to CoreML, Apple's ML framework, where it is classified in the app using YOLO, which is one of several models that can be integrated into CoreML. As an example for indoor scenes, five new classes were implemented, namely tvmonitor, laptop, bed, sink and toilet. The integration of YOLO was a particular challenge in terms of software technology, as mesh classification with ARKit and object detection with YOLO could not be processed in the same procedure. Instead, the processes were split, first the classification with ARKit and then detection with YOLO. The processes could also not be parallelised in the prototype because of the permanently regeneration of the mesh geometry by ARKit. That is why YOLO processes static areas after they are already classified as *none*.

3.3 AR-based Interactive Visualisation

To visualise the detection process on the mobile device, the classified mesh is overlayed with the real camera image, i.e. AR is created. In case of successful object detection with both the eight ARKit standard classes and the extended five YOLO classes, the affected area of the mesh is coloured associated to the object class. The SceneKit framework [20] (among others) was used for visualisation on the mobile device screen in real-time. SceneKit displays a 3D scene, such as the virtual image created while capturing the interior spaces, on the screen. It calculates which elements of the generated mesh are visible from the current camera angle and displays them on the screen. Since the colouring of the mesh runs parallel to the classification, the mesh generated

by ARKit is coloured first and then, after the scan has been completed, all grey (i.e. ARKit class *none*) mesh parts are processed with YOLO and coloured in terms of highest class probability accordingly to the object detected. For this purpose, the objects detected in the camera frames by YOLO,as explained in the previous section are then spatially assigned and reprojected from the 2D screen to all anchors in the 3D mesh by raycasting operations. The result can be seen in the following Fig. 2, where an object of the class *laptop* is classified first as *none* (c.f. Fig. 2a). After that, the region is detected by YOLO, thus framed in yellow and reprojected to the generated anchors of the 3D mesh by ARKit (c.f. Fig. 2b).



(a) (b)

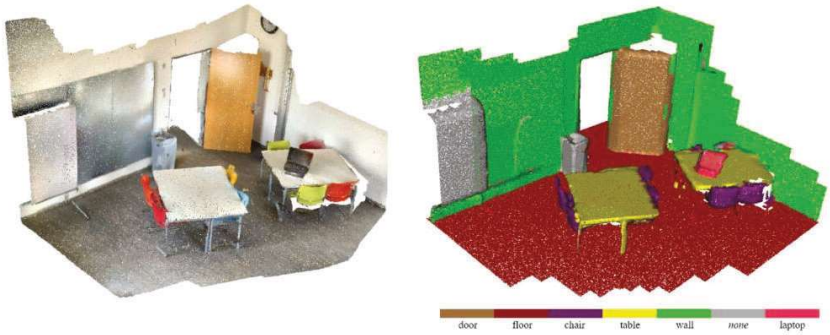
Fig. 2. Acquisition process of a specific object that is classified first as *none* with ARMeshClassification (a) and later finalised with YOLO as *laptop* (b). The yellow bounding box represents the region of interest for YOLO that is projected to the captured mesh surface by ray traces. The video sequence[5] according to that figure was captured by an Apple iPad Pro in debug mode with 30 fps.

4 Result and Outlook

This paper presented the mobile app "Semantic Data Capture" based on iOS for (a) capturing 3D building sites with consumer products and (b) recognising interior structures and furnishings using efficiently augmented reality (AR) and machine learning (ML) methods. This serves to support the AEC industry in general for planning purposes and facility management for e.g. automated space analysis with respect to BIM. It was shown how user-centred working can also be made possible for non-expert users with the help of an integrated LiDAR sensor and augmented reality. For this, consumer hardware from Apple was used, which is currently the only manufacturer offering the functionality of active low-cost and real-time 3D scanning. With a combination of both, in-built sensors and close software libraries using AR and ML techniques, a new method for Scan-to-BIM was suggested and successfully prototyped. Here, Apple's own

[5] Video sequence of "Semantic Data Capture": https://video.hs-mainz.de/Panopto/Pages/Viewer.aspx?id=6838959a-83af-4e36-918d-ad970123048d

recognition methodology for furnishings was extended with classes from the freely available YOLO model. The results after the data export are shown in Fig. 3.



(a) (b)

Fig. 3. Results of the app "Semantic Data Capture" (b) illustrated by CloudCompare in comparison to an RGB coloured point cloud captured by 3D Scanner App[1] (a). The first six classes are from the ARKit mesh classification while the class *laptop* originated from the extended YOLO model.

Despite the successful implementation of the prototype, there are several challenges that need to be considered in future. First, the simultaneous mesh classification by ARKit and object detection by YOLO could not yet be implemented. This is one of the improvement possibilities envisaged for the immediate future. Improving this aspect, the dual data acquisition could be avoided and working in real-time will be possible. Further additions can be made in the area of object detection with ML models. For example, YOLO or other models can be used to add further classes for additional components or furnishings at free will. As another main improvement, the acquired data could be used in combination with existing models to generate more precise ones individually prepared for the situation through training. One of the core disadvantages of the solution presented here is the utilisation of libraries provided by Apple that cannot be viewed and changed by their black box character. Considering the geometrical accuracy of the captured and classified objects, LiDAR could not keep up terrestrial laserscanning or SfM techniques in terms of reliability and precision. As a consequence, LiDAR results serves currently just for coarse BIM modelling.

Nevertheless, in future, the ability of LiDAR will increase with respect to acquisition and registration accuracy. By means of that, LiDAR will be among the most important acquisition methods for 3D point cloud capturing even for fine geometry and BIM valid model fitting. In collaboration with powerful and smart mobile devices, the possibilities to process 3D data are so far not limited as the prototype shows. "Semantic Data Capture" provides a new paradigm of building indoor acquisition and processing simultaneously in order to achieve the final scope of Scan-to-BIM: the automated 3D modelling of buildings with appropriate accuracy by using mobile devices.

5 Acknowledgements

The authors would like to thank D. Iordanov for the contribution to the development of the application within his studies at Mainz University of Applied Sciences.

References

1. López Iglesias, J.; Díaz Severiano, J. A.; Lizcano Amorocho, P.E.; Del Manchado Val, C.; Gómez-Jáuregui, V.; Fernández García, O. et al. (2020): Revision of Automation Methods for Scan to BIM. In: Advances in Design Engineering. Cham: Springer International Publishing (Lecture Notes in Mechanical Engineering), pp. 482–490.
2. Plaß, B.; Emrich, J.; Goetz, S.; Kernstock, D.; Klauer, T. (2021): Evaluation of point cloud data acquisition techniques for Scan-to-BIM workflows in Healthcare. In: Proceedings of the FIG e-Working Week 2021. Amsterdam.
3. Plaß, B.; Prudhomme, C.; Ponciano, J.J. (2021): BIM ON ARTIFICIAL INTELLIGENCE FOR DECISION SUPPORT IN E-HEALTH. In: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B2-2021, pp. 207–214. DOI: 10.5194/isprs-archives-XLIII-B2-2021-207-2021.
4. Tang, P.; Huber, D.; Akinci, B.; Lipman, R.; Lytle, A. (2010): Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. In: Automation in Construction 19 (7), pp. 829–843. DOI: 10.1016/j.autcon.2010.06.007.
5. Volk, R.; Stengel, J.; Schultmann, F. (2014): Building Information Modeling (BIM) for existing buildings – Literature review and future needs. In: Automation in Construction 38, p. 109–127. DOI: 10.1016/j.autcon.2013.10.023.
6. Loges, S.; Blankenbach, J. (2017): As-built Dokumentation für BIM - Ableitung von bauteilorientierten Modellen aus Punktwolken. Photogrammetrie - Laserscanning - optische 3D-Messtechnik. In: Beiträge der Oldenburger 3D-Tage, pp. 290–298.
7. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. (2017): PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 77-85.
8. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. (2017): PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 5105-5114.
9. Li, Y.; Bu, R.; Sun, M.; Wu, W.; Di, X.; Chen, B. (2018): PointCNN: Convolution on x-transformed points. In: Advances in neural information processing systems, pp. 820–830.
10. Hu, Q.; Yang, B.; Xie, L.; Rosa, S.; Guo, Y.; Wang, Z.; Trigoni, N.; Markham, A. (2020): RandLA-Net: Efficient semantic segmentation of large-scale point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11108–11117.
11. Huang, Q.; Wang, W.; Neumann, U. (2018): Recurrent slice networks for 3D segmentation of point clouds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2626–2635.
12. Phan, A.V.; Le Nguyen, M.; Nguyen, Y.L.H.; Bui, L.T. (2018): DGCNN: A convolutional neural network over largescale labeled graphs. In: Neural Networks (108), pp. 533–543.
13. Berger, M.; Tagliasacchi, A.; Seversky, L.; Alliez, P.; Guennebaud, G., Levine, J.; Sharf, A.; Silva, C. (2016): A Survey of Surface Reconstruction from Point Clouds. In: Computer Graphics 36 (1), pp. 301-329. DOI: 10.1111/cgf.12802.
14. Xiaoyi, R.; Baolong, L. (2020): Review of 3D Point Cloud Data Segmentation Methods. In: International Journal of Advanced Network, Monitoring and Controls 5 (1), pp. 66–71. DOI: 10.21307/ijanmc-2020-010.

15. Ponciano, J.J.; Roetner, M.; Reiterer, A.; Boochs, F. (2021): Object Semantic Segmentation in Point Clouds – Comparison of a Deep Learning and a Knowledge-Based Method. In: ISPRS Int. J. Geo-Inf. 10 (4). DOI: 10.3390/ijgi10040256.
16. Apple ARKit Documentation (2021): https://developer.apple.com/documentation/arkit, accessed on Sept., 1., 2021.
17. Apple ARMeshClassification Documentation (2021): https://developer.apple.com/documentation/arkit/armeshclassification, accessed on Sept., 1., 2021.
18. Redmon, J; Divvala, S.; Girshick, R.; Farhadi, A. (2016): You Only Look Once: Unified, Real-Time Object Detection. In: Proceedings Conference on Computer Vision and Pattern Recognition, arXiv:1506.02640.
19. Lin, T.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollar, P.; Zitnick, C. L. (2014): Microsoft COCO: Common Objects in Context. In: Computer Vision -- ECCV 2014, European Conference on Computer Vision, Springer International Publishing 2014, pp.740-755.
20. Apple SceneKit Documentation (2021): https://developer.apple.com/documentation/scenekit/ , accessed on Sept., 1., 2021.