

INDOOR MESH CLASSIFICATION FOR BIM

L. S. Runceanu^{1,*}, N. Haala¹

¹ Institute for Photogrammetry, University of Stuttgart, Germany – (lavinia.runceanu, norbert.haala)@ifp.uni-stuttgart.de

Commission IV, WG IV/1

KEY WORDS: Mesh classification, BIM, Random Forest

ABSTRACT:

This work addresses the automatic reconstruction of objects useful for BIM, like walls, floors and ceilings, from meshed and textured mapped 3D point clouds of indoor scenes. For this reason, we focus on the semantic segmentation of 3D indoor meshes as the initial step for the automatic generation of BIM models. Our investigations are based on the benchmark dataset ScanNet, which aims at the interpretation of 3D indoor scenes. For this purpose it provides 3D meshed representations as collected from low cost range cameras. In our opinion such RGB-D data has a great potential for the automated reconstruction of BIM objects.

1. INTRODUCTION

Building Information Modelling, which became popular starting with 2002, is considered to be an intelligent 3D model which focuses on design, construction and management of a building site (Autodesk, 2018). Even though the use of Building Information Models (BIM's) for the current building constructions is state of the art, there is still ongoing research to automatically create these models starting from the scanning data and to keep them updated in time. Also a great challenge is to automatically deliver BIM's for older buildings which do not dispose of as-designed models. Due to the fact that a considerable big number of buildings are in the last-mentioned situation, there is a need of automated reconstruction methods which can deliver from input data suitable models that enables the creation of BIM objects. Currently the Industry Foundation Classes (IFC) format is used as an open standard format for the BIM objects. According to (Ohuri et al., 2017), most IFC objects are built by integrating sweep volumes, explicit faceted surface models and Constructive Solid Geometry (CSG).

For obtaining the needed models a lot of indoor data needs to be acquired. According to (Runceanu et al., 2017) active systems are often use for mapping indoor environments, overcoming very good the lack of texture problem. Even though, traditionally, a laser scanner is used for acquiring the indoor data, recently a larger variety of sensors is available. Indoor Mobile Mapping Systems (IMMS), like M6 trolley from NavVis (NavVis M6, 2018) integrate different sensors and algorithms, in order to reduce the mapping cost and increase the efficiency. Also, sensors like Microsoft Kinect (MSDN Kinect, 2018), DPI-8 (DotProduct DPI-8, 2018) and Google Tango tablet (Google Tango, 2018) made the interest growing in using low-cost range camera sensors for mapping. However most of these new equipment lack in the delivered accuracy. For this reason, appears the question if a low cost system, integrating range cameras, can deliver data accurate enough for creating BIM objects. At this question we aim to answer.

The 3D data format, delivered by these sensors, differ. Even if point cloud data still remains a standard format for these kind of tasks, also 3D meshes and voxel grids seem to be more and

more often used as input data, both in real applications and for research purposes. Having the purpose of reconstructing BIM objects, we considered that voxel grids have computational advantages, but they lack in modelling accuracy. However, voxel grids enable the use of a 3D deep network for classifying the 3D data. On the other side, meshes, compared to point clouds, have the advantages of building closed surfaces, containing image texture and facilitating the normal and neighbourhood computation. This motivated us to use the mesh format as an input format. Consequently this type of data needs to be semantically interpreted and modelled in order to be later integrated in the compatible BIM format.

All these aspects motivated us to make use of an existing indoor benchmark containing indoor data coming from a low cost sensor in order to classify indoor environments as an important step in BIM creation. More specifically we use the ScanNet indoor benchmark (Dai et al., 2017) in order to train a classifier. From the classified classes we mainly focus on three of them: walls, floors and ceilings, all the others being filter out.

The RGB-D data from this benchmark was acquired by the structure sensor from Occipital. The sensor has the possibility of measuring the distance to the surrounding objects in a range of 0.3 – 3.5 m, with an accuracy varying from 0.1-1.1% (Occipital Structure, 2018). The advantages of this benchmark dataset are on one side the big size of the data, enabling different test scenarios and on the other side that it provides raw RGB-D data and the camera poses which enables a volumetric fusion (Curless and Levoy, 1996) and the extraction of the surface mesh. Considering the aforementioned advantages of mesh format, a possibility will be to classify directly the 3D data, but also to integrate classification results of the raw RGB-D data. The last option is subject of further work, therefore this current work focuses on mesh classification. According to (Dai et al., 2016) the surface reconstruction accuracy is below 1 cm. However, an improvement of the mesh accuracy was not subject of this work.

Our algorithm consists of the mesh patches generation by a region growing segmentation and then the classification of the resulted segments by using a Random Forest algorithm. This

* Corresponding author

work is structured as follows. Section 2 presents an overview of the related work focused on the mesh segmentation techniques and on the automation of the scan-to-BIM process. The methodology is presented in section 3. The experimental results are given in section 4. Conclusions are given in section 5.

2. RELATED WORK

The related work is structured considering the mesh segmentation techniques and the automation of the scan-to-BIM process.

2.1 Mesh segmentation techniques

Extensive research work in the field of 3D city modeling proved that semantic interpretation of a scene is very important for an accurate 3D reconstruction (Riemenschneider et al., 2014; Martinović et al., 2015; Bláha et al., 2017). Also, mesh segmentation techniques proved to be useful in the 3D reconstruction process for indoor and outdoor scenarios. (Kähler and Reid, 2013) classified indoor environments coming from RGB-D images by using Decision Tree Fields (Nowozin et al., 2011) and Regression Tree Fields (Jancsary et al., 2012). They started with a dense 3D reconstruction and after that they performed an oversegmentation, inspired by the SLIC superpixel algorithm. (Valentin et al., 2013) presented their own approach of building a triangulated mesh representation from multiple depth estimates. They used a CRF approach and in this framework they were able to consider both the geometric properties coming from the 3D mesh and the visual ones coming from the RGB-D images. (Dai et al., 2017) created a dataset of annotated RGB-D scans of indoor environments, containing 2.5M of RGB-D images. Using this dataset it was possible to train a 3D deep network and perform several scene understanding tasks, like 3D object classification, semantic voxel labelling, and CAD model retrieval. Due to the structure of the 3D neural network the meshes were not directly classified. We considered that the mesh structure, which is keeping the topology, is more suitable for the later conversion to BIM objects. This is why, in this work, the 3D data as a mesh is passed through a classification pipeline. The pipeline is inspired by the semantic segmentation algorithm for urban scenes, proposed by (Rouhani et al., 2017).

2.2 Scan-to-BIM process

Various works address the problem of automatizing the process of converting input scanned data into a BIM, which mainly consist of preprocessing the data, including also the choice of a suitable format, then segmentation and classification of the determined segments and parameters extraction for the BIM reconstruction. (Xiong et al., 2013) aimed at modelling the main structural components of indoors, like: walls, floors, ceilings, windows and doors. They also addressed in their work the challenges of clutter (confusion) and occlusion by explicitly reasoning about them through the process. Their algorithm operated only on planar patches and automatically learned features and contextual relationships from training data. Main failures occurred in the interiors of low built-in cabinets and stairwells. (Tuttas et al., 2014) used point clouds delivered from unordered images in order to monitor a construction progress. They also performed a comparison between the as-planned and as-built states of the construction with the help of an octree-structure. However, a prior for this work is the need for intermediate monitoring data in order to constantly update the model. (Bassier et al., 2018) presented a method to automatically

reconstruct wall geometry from point clouds in a BIM standard format. Their method is suitable for complex, multi-storey buildings. However they made some assumptions, i.e. the floors and the ceilings are planar and also working with almost complete point clouds, reduced the number of challenges. (Macher et al., 2017) proposed a semi-automatic approach for 3D reconstruction of indoors from point clouds. Walls and slabs of the building were reconstructed in the Industry Foundation Classes (IFC) standard BIM format. The last two aforementioned works motivated us to try to obtain similar results, but from low cost meshed data.

3. METHODOLOGY

The goal of our approach is to detect and reconstruct BIM objects from input meshed point cloud data. Firstly we focus on the following BIM objects: walls, floors, ceilings. Later this work will be extended to reconstruct openings, like doors and windows and also furniture.

In order to implement the proposed algorithm, meshes from the ScanNet benchmark were used (Dai et al., 2017). Initially the meshes were oversegmented with the help of a region growing segmentation algorithm, which divided them into small patches (Figure 3). This algorithm was inspired by the work of (Rouhani et al., 2017), who also computed “superfacets” before classifying the mesh. The process starts by picking a random face as a seed. For every seed-face it is computed the normal and the mean colour in HSV space. Starting from the gravity centre of each face, a spherical neighbourhood is considered with a given radius. The faces with all the vertices inside this neighbourhood are further considered in the processing.

$$N = \{f \in F \mid \|f - s\| < d\} \quad (1)$$

where N = neighbourhood
 f = face
 F = all the faces
 s = seed face
 d = search radius

The criteria, on which it is considered that different neighbouring faces belong to the same region, it is the colour similarity and the normal orientation. The colour similarity is defined as the L1 distance in the HSV colour space between the mean colour of the seed face and the mean colour of the respective face. If this distance is smaller than a threshold, then it is also computed the angle between the seed normal and the normal of the neighbouring face. If this angle is smaller than a threshold, then it is considered that the neighbouring face is belonging to the same region.

$$R = \left\{ n \in N \mid \|hsv(n) - hsv(s)\| < th_{hsv} \wedge (\vec{n}_n \cdot \vec{n}_s) > th_{angle} \right\} \quad (2)$$

where n = neighbouring face
 $hsv()$ = mean colour in HSV space of a face
 th_{hsv} = colour threshold
 th_{angle} = angle threshold
 \vec{n}_n = normal of the neighbouring face
 \vec{n}_s = normal of the seed face

The process is repeated iteratively and finally it will result a set of regions, containing mesh faces with similar properties.

The oversegmented meshes, chosen to be part of a training set, were used to train a Random Forest classifier and the learned knowledge was then used to predict the classes of other test meshed point clouds. For each resulted patch from the oversegmentation process, geometric and radiometric features were computed. (Rouhani et al, 2017) and (Kähler and Reid, 2013) provided a detailed list of the used features for mesh classification. Their suggestions inspired also our feature selection. As a first geometric feature, the mean of all face normals of a patch is considered. The next geometric feature is computed as the cosine of the angle between the mean normal of the patch and the vertical axis. This features is measuring the verticality of the patch being useful to differentiate between vertical and horizontal objects. Also, the mean height of the patch is a next important geometric feature. It is computed as a mean of all the vertices' heights from the patch, with the reference at the floor level. Therefore it helps differentiating between similar indoor objects, which are though always at different height located, like ceiling and floor. The radiometric information completes the geometric one in classifying the patches. For each patch it was computed the mean colour and its corresponding standard deviation in the HSV colour space. For training a random forest approach all these features were concatenated in a feature vector. The results of the classification are used on the one side to filter out the classes which, for the moment, are not considered to be objects of interest, like plants and furniture. On the other side the remained classified objects are used to extract the parameters for the class-characteristic 3D reconstruction. For the class walls, the normal orientation is useful for fitting planes to each individual wall. If neighbouring rooms are available also the wall thickness can be extracted. This will enable a conversion from a surface to a solid entity. An option in this regard will be the use of the open-source FreeCAD software (FreeCAD, 2018), which allows a direct conversion into the IFC format.

4. EXPERIMENTS

Because this work is mainly focused on public buildings, first tests were performed using meshes from the office category (Figure 1). The mesh oversegmentation was realised considering a search radius of 0.1 m for the neighbourhood (Figure 2). The colour threshold was set to $th_{hsv} = 20$ and the angle threshold to $th_{angle} = 15^\circ$. Working with indoor environments where the majority of object classes are planar, it proved to be very useful to perform first an oversegmentation that forced to consider also the mesh face neighbourhood (Figure 3 in comparison with Figure 4). In our Random Forest implementation we used 50 classification trees. Figure 5 shows the prediction made by the classifier for a test room (Figure 4). By comparing the classification result with the true labels (Figure 5), it results that some classes, like the floors are completely detected, while others include outliers. For the class walls it seems that similar planar objects, like part of a shelf and table are misclassified as walls. In order to overcome this issue, on the one side, the connected patches classified as walls below a threshold could be filter out. On the other side, it is planned to further use the second version of the training benchmark (ScanNet v2, 2018), which was made available during writing of this work. The new version of the benchmark increases the labelled surface coverage from 63% to 90%, which will allow that the objects which were unlabelled to contribute to the classification and to be correct classified.



Figure 1. Sample mesh

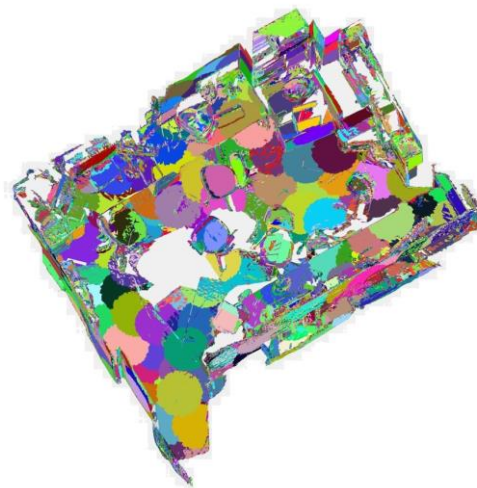


Figure 2. Oversegmentation of the sample mesh

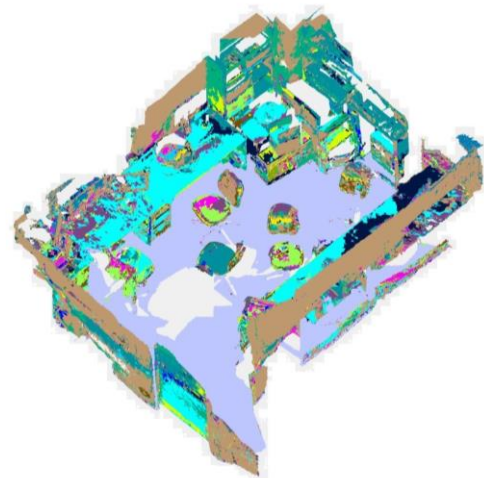


Figure 3. Random Forest classification results (without oversegmentation)

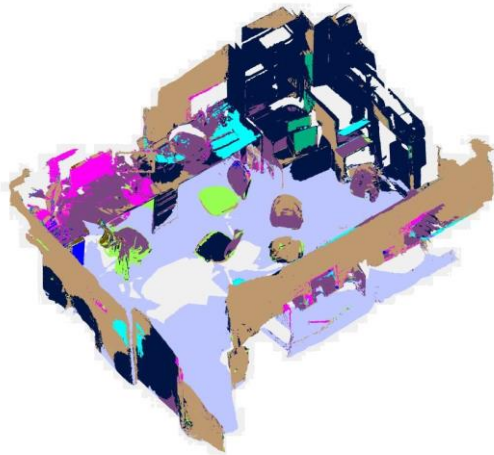
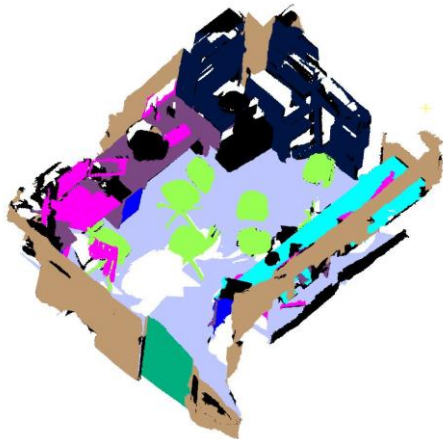


Figure 4. Random Forest classification results (after oversegmentation)



wall	chair	floor	table	door	shelf	desk	monitor
computer	drawer	keyboard	computer towerspeaker	jacket	plug/outlet	unannotated	

Figure 5. True labels from the benchmark ScanNet

5. CONCLUSIONS

This paper presents a mesh classification algorithm for low-cost 3D data with the purpose of detecting classes useful for BIM object creation. The proposed algorithm was tested on the ScanNet benchmark dataset and proved to deliver good results, enabling us to further work on the 3D reconstruction for the aforementioned objects of interest. Some of the challenges that appeared, in the form of misclassification, are expected to be overcome by using a newer version of the benchmark dataset and by realizing an integration of this 3D mesh classification with a RGB-D classification, the needed 2D semantic labels

being provided by the benchmark. Another improvement for the different classes of furniture, which will also be of interest in the future, will be realized by merging some similar classes of the training dataset into one class. As an example “table” and “desk” could be just one class.

REFERENCES

- Autodesk, 2018. <https://www.autodesk.com/solutions/bim> (11 July 2018).
- Bassier, M., Klein, R., Van Genechten, B., and Vergauwen, M., 2018. IFC Wall reconstruction from unstructured point clouds. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. IV-2, pp. 33-39, <https://doi.org/10.5194/isprs-annals-IV-2-33-2018>.
- Bláha, M., Rothermel, M., Oswald, M.R., Sattler, T., Richard, A., Wegner, J.D., Pollefeys, M., and Schindler, K., 2017. Semantically informed multiview surface refinement. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, pp. 3839-3847.
- Curless B. and Levoy M., 1996. A volumetric method for building complex models from range images. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM, pp 303–312.
- Dai, A., Nießner, M., Zollöfer, M., Izadi, S., and Theobalt, C., 2016. BundleFusion: Real-time globally consistent 3D reconstruction using on-the-fly surface re-integration. *arXiv preprint arXiv:1604.01093*.
- Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., and Nießner, M., 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (Vol. 1).
- DotProduct DPI-8, 2018. <https://www.dotproduct3d.com/dpi8x.html> (12 July 2018).
- FreeCAD, 2018. <https://www.freecadweb.org/> (12 July 2018)
- Google Tango, 2018. [https://en.wikipedia.org/wiki/Tango_\(platform\)](https://en.wikipedia.org/wiki/Tango_(platform)) (12 July 2018).
- Jancsary, J., Nowozin, S., Sharp, T. and Rother, C., 2012. Regression Tree Fields—An efficient, non-parametric approach to image labeling problems. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, IEEE, pp. 2376-2383.
- Kähler, O. and Reid, I., 2013. Efficient 3d scene labeling using fields of trees. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3064-3071.
- Macher, H., Landes, T. and Grussenmeyer, P., 2017. From Point Clouds to Building Information Models: 3D Semi-Automatic Reconstruction of Indoors of Existing Buildings. *Applied Sciences*, 7(10), pp.1030.
- Martinovic, A., Knopp, J., Riemenschneider, H., and Van Gool, L., 2015. 3d all the way: Semantic segmentation of urban scenes from start to end in 3d. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4456-4465.

MSDN Kinect, 2018. <https://msdn.microsoft.com/en-us/library/hh438998.aspx> (12 July 2018).

NavVis M6, 2018. <https://www.navvis.com/m6> (12 July 2018).

Nowozin, S., Rother, C., Bagon, S., Sharp, T., Yao, B., and Kohli, P., 2011. Decision tree fields. In: *Computer Vision (ICCV), 2011 IEEE International Conference on*, IEEE, pp. 1668-1675.

Occipital Structure, 2018. <https://structure.io/> (12 July 2018)

Ohuri, K. A., Biljecki, F., Diakit , A., Krijnen, T., Ledoux, H., and Stoter, J., 2017. Towards an integration of GIS and BIM data: What are the geometric and topological issues. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. IV-4/W5, pp. 26-27, <https://doi.org/10.5194/isprs-annals-IV-4-W5-1-2017>.

Riemenschneider, H., B dis-Szomor , A., Weissenberg, J., and Van Gool, L., 2014. Learning where to classify in multi-view semantic segmentation. In: *European Conference on Computer Vision*, Springer, Cham, pp. 516-532.

Rouhani, M., Lafarge, F., Alliez, P., 2017. Semantic Segmentation of 3D Textured Meshes for Urban scene Analysis. In: *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, 123, pp. 124-139. <10.1016/j.isprsjprs.2016.12.001>. <hal-01469502>.

Runceanu, L. S., Becker, S., Haala, N., and Fritsch, D., 2017. Indoor Point Cloud Segmentation for Automatic Object Interpretation. In: *Proc. of the 37. Wissenschaftlich-Technische Jahrestagung der DGPF*.

ScanNet v2, 2018. <http://www.scan-net.org/changelog#scannet-v2-2018-06-11> (12 July 2018).

Tuttas, S., Braun, A., Borrmann, A., and Stilla, U., 2014. COMPARISON OF PHOTOGRAMMETRIC POINT CLOUDS WITH BIM BUILDING ELEMENTS FOR CONSTRUCTION PROGRESS MONITORING. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XL-3.

Valentin, J.P., Sengupta, S., Warrell, J., Shahrokni, A., and Torr, P.H., 2013. Mesh based semantic modelling for indoor and outdoor scenes. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2067-2074.

Xiong, X., Adan, A., Akinci, B., and Huber, D., 2013. Automatic creation of semantically rich 3D building models from laser scanner data. *Automation in Construction*, 31, pp. 325-337.