ONTOLOGY-BASED SEMANTIC LABELING FOR RGB-D AND POINT CLOUD DATASETS

Fabian Kaufmann[1], Mahdi Chamseddine[1,2], Suresh Guttikonda[2]
Christian Glock[1], Didier Stricker[1,2], Jason Rambach[2]
[1]RPTU Kaiserslautern-Landau, Kaiserslautern, Germany
[2]German Research Centre for Artificial Intelligence, Kaiserslautern, Germany

Abstract

Applications of deep learning have recently seen a surge in the field of construction. Supervised semantic segmentation of 2D or 3D data acquired from buildings requires the use of annotated data for training, validation, and testing. Although various datasets have been published targeting this application, they lack a common convention and definitions based on construction ontologies. This work presents a guideline for ontology-based semantic annotation of RGB-D and point cloud datasets for buildings. Such a contribution facilitates the use of deep learning in construction by bridging the gap between this field and computer science.

The annotation guideline is available under this link `https://gitlab.rhrk.uni-kl.de/kaufmann/humantech-data-annotation`.

Introduction

Several techniques are available to map and measure 3D spaces, namely LiDAR and RGB-D sensors. LiDARs (Light Detection and Ranging) are time-of-flight (ToF) devices that measure the distance and reflectivity of objects by emitting light and measuring the time it takes for it to be reflected back. Each emitted and reflected ray by the LiDAR corresponds to one measurement point and thousands of such measurements constitute a 3D point cloud. On the other hand, RGB-D sensors combine a color camera (RGB) with a distance measurement technique (D) such as: structured light (Geng, 2011), time-of-flight, or stereo vision, which is the use of two cameras to infer the geometry through difference between the image perspectives. The measurement techniques have their strengths and weaknesses, however, RGB-D has the added benefit of not only capturing depth but also color information.

To facilitate the integration of data into state-of-the-art BIM workflows, the acquired RGB-D or 3D point cloud data must be translated into a BIM automatically (Kaufmann et al., 2022). One important component of scan-to-BIM automation is the process of semantic segmentation i.e., assigning a class to every point of the point cloud. A wide range of approaches have been investigated in this field. The progress can be studied following the recent submissions to benchmark challenges such as ScanNet (Dai et al., 2017). All approaches relying on supervised learning require data with annotations. Although various datasets exist, there are no conventions on the ontologies and classes used, and, more importantly, no annotation rules have been defined to ensure consistent annotation among datasets.

Data annotation is also typically performed by multiple persons. If only the labels without any further description and annotation guidelines are presented, it is likely that annotation will be performed differently based on the experience of the persons involved. A civil engineer will have a deeper understanding of the construction ontology from his experience in the field and thus make other decisions than a person not from the construction domain. This problem cannot be fully overcome using guidelines. The number of false positives can be reduced thus leading to better training results.

The objective of this paper is to propose unified labeling guidelines for RGB-D and point cloud annotation based on commonly used construction ontologies. This not only has the potential to serve as a basis for consistent data annotation strategies, but also opens the door to more systematic and reliable evaluation of deep learning algorithms and open the door for training on multiple datasets.

At the current state, the guidelines are applicable for buildings. Beside components of the building itself, typical labels for objects on construction sites (scaffolding, formwork) are defined and interior objects such as furniture are covered as well. So the annotation guideline can be applied for data from existing buildings, buildings under construction and buildings under operation.

Related Work

A large number of RGB-D and point cloud datasets have been captured for indoor scenes, which are used for computer vision tasks: 3D scene understanding (Gupta et al., 2013), semantic segmentation (Gupta et al., 2014; Ren et al., 2012), 3D object detection (Lin et al., 2013), and others.

Silberman et al. (2012) collected a wide range of commercial and residential buildings, comprising 464 different indoor scenes of offices, stores, and rooms of houses. The resulting dataset is composed of $407,024$ RGB-D images. Among these dataset images, 1449 image frames are labeled, covering 894 object classes categorized into 40 object categories.

Song et al. (2015) gathered a dataset from four RGB-D sensors: the Intel RealSense 3D Camera, the Asus Xtion LIVE PRO, and the Microsoft Kinect V1 and V2. This produced $10,335$ RGB-D indoor images that are extensively annotated using 2D polygon and 3D bounding box annotations in the style of Russell et al. (2005).

The 2D-3D dataset introduced by Armeni et al. (2017) comprises RGB, depth, equirectangular, global XYZ, 3D meshes, and point clouds (Armeni et al., 2016) of the indoor spaces. The dataset is gathered in six large-scale indoor areas of three separate educational and office buildings. Pixel-to-pixel correspondence exists between all modalities across 13 object classes in the collected $70,496$ RGB images.

In a style analogous to Armeni et al. (2017), Chang et al. (2017) used Matterport cameras to acquire a 3D dataset. Data is nonetheless collected from 90 different buildings that contain a large variety of scenarios: residences, offices, and churches. In contrast to Armeni et al. (2017), this work includes 40 object categories, providing good coverage of both building elements and objects.

Dai et al. (2017) crowd-sourced the manual labeling work of 2.5 million RGB-D image views captured from a handheld device. Meanwhile, Zheng et al. (2020) present a large-scale photo-realistic synthetic dataset of 3D structure and photo-realistic 2D renderings of indoor man-made environments. As a result, the laborious process of manual labeling has been avoided in this work. The dataset comprises $196,515$ images annotated into 40 object categories, similarly as Silberman et al. (2012).

While existing datasets have been useful for scene understanding and other computer vision tasks, they fall short of fulfilling the requirements of the construction domain. Although annotations in existing datasets cover common construction classes such as doors, windows, floors, and ceilings, they do not include other structural components, such as slabs, columns and beams and elements involved in construction itself such as scaffolding or formwork, to name a few. Furthermore, to use the data for applications such as scan-to-BIM, annotations based on construction ontologies are necessary. Finally, different datasets follow their own set of guidelines, making it challenging to evaluate algorithms and compare their performance on different data.

To address these shortcomings of existing datasets, an annotation guideline is needed that covers both image and point cloud modalities, and includes proper instructions for labeling the desired construction classes accurately, as shown in Table 1. This will ensure uniform quality among different construction-oriented datasets.

Ontologies in Construction

An ontology aims to describe the nature of physical objects and their relationship to each other. In construction, ontologies describe the structure of assets of the built environment, their objects, components and abstract entities (e.g., actors, resources). Ontologies can also be described as hierarchical classification schemes which make them a valuable source of knowledge for the classification and annotation of data from buildings such as RGB-D images or point clouds. Related to construction, different ontologies exist, some of which are related to a data model such as Industry Foundation classes (IFC). (Buildingsmart, 2020).

Although different ontologies such as Uniclass, a five hierarchy classification scheme provided by the National Buildings Specification (NBS) (National Building Specification, 2022), and OmniClass, a similar classification schema for the US construction industry with a more specific focus on the entire building life cycle (Construction Specification Institute, 2023) exist, we will comply to the IFC standard in this work. Not only does IFC contain basic entity definitions, it also comes with a data model. Data annotated based on the IFC ontology, and algorithms trained with such data, facilitates the use of the algorithms for various BIM use cases e.g., generating as-built models, progress monitoring, construction site asset tracking and others.

Although there are entity definitions in IFC, a detailed description of the entity or class is not provided. As an example, the entity *IfcWall* is defined as follows: *"The wall represents a vertical construction that bounds or subdivides spaces. Walls are usually vertical, or nearly vertical, planar elements, often designed to bear structural loads. A wall is however not required to be load bearing"* (Buildingsmart, 2020).

However, a more detailed description is crucial for data annotation. Such a description should comprise a definition of what objects, assets or other representations are in the class. Additionally, the topology and boundaries with other objects as well as annotation rules should be defined.

A comprehensive description of entities in the context of data annotation seems trivial. In most cases it is not, as the example of walls and slabs will demonstrate. Seen from the inside, a wall is limited by the ceiling or floor, or the slabs if the structure is visible. From the outside of a building where floor, ceiling and slabs are typically not visible, a rule needs to be defined how and where wall instances would be considered as separate elements. Balconies could be considered as part of the slab, when the monolithic slab is extended to the outside, or as a separate class. It becomes obvious, that rules besides the existing

Table 1: Comparison of object categories across different datasets and our proposed ontology-based categorization.

	Scene	Object Categories
Silberman et al. (2012) Chang et al. (2017) Dai et al. (2017) Zheng et al. (2020)	Indoor	void, wall, floor, cabinet, bed, chair, sofa, table, door, window, bookshelf, picture, counter, blinds, desk, shelves, curtain, dresser, pillow, mirror, floor mat, clothes, ceiling, books, refrigerator, television, paper, towel, shower curtain, box, whiteboard, person, night stand, toilet, sink, lamp, bathtub, bag, other-structure, other-furniture, and other-prop
Song et al. (2015)	Indoor	void, wall, floor, cabinet, bed, chair, sofa, table, door, window, bookshelf, picture, counter, blinds, desk, shelves, curtain, dresser, pillow, mirror, floor mat, clothes, ceiling, books, fridge, tv, paper, towel, shower curtain, box, whiteboard, person, night stand, toilet, sink, lamp, bathtub, and bag
Armeni et al. (2017)	Indoor	unknown, beam, board, bookcase, ceiling, chair, clutter, column, door, floor, sofa, table, wall, and window
OURS	Building	slab, floor, ceiling, wall, pipe horizontal, pipe vertical, fitting, door, window, stair, roof, column, beam, truss, chimney, railing, ramp, elevator, pavement, gravel pad, and curtain wall
	Construction	scaffolding and formwork
	Interior	invalid, lamp, cctv camera, shelf, box, trash bin, first aid kit, fire extinguisher, exit sign, computer hardware, table, plant pot, picture, cabinet, chair, and couch

ontologies need to be defined. In the next section a set of classes with annotation rules is presented. Wherever possible, classes are directly related to IFC classes and the IFC entity definitions are extended with further descriptions and annotation rules.

Annotation Guidelines

The annotation guidelines are designed to serve as a common ground for labeling both point clouds and RGB-D data, so special care was taken while defining the different classes and their annotation guidelines. As shown in Table 1, three different categories were observed when defining the labeled classes:

- Building Category: this includes the subset of labels that belong to the building itself, such as walls, doors, columns, windows, etc. All items under the building category are fixed to the building.
- Construction Category: this category includes objects used in construction, such as formwork and scaffolding.
- Interior Category: includes a large set of labels that can be seen inside a building or on a construction site, such as tables, chairs, computer hardware, fire extinguishers, etc. Interior objects are movable and thus not fixed to the building.

In addition to those categories, some special labels were defined for useful categories such as invalid data for over and underexposed images where labeling is not possible as in Fig. 1 (left) or sky where the depth data is invalid but the image can be clearly identified as sky. In point clouds, clutter, noise, and unidentifiable artifacts should be annotated as invalid data as seen in Fig. 1 (right). Those labels are important when training and validating neural networks to reduce false positives and are required for real-scene data.

The guidelines introduced for annotation differ from existing datasets by conforming to the IFC classes whenever possible and creating a common framework for both 3D (point clouds) and 2D (RGB images and RGB-D images) labels. This guideline can thus be used for labeling data for different stages of construction and making it possible to generate BIM models from the data. Additionally, a consistent annotation of 2D and 3D data will facilitate the data comparison on a semantic level and thus allow us to investigate the specific advantages of different sensors.

From the 43 labels defined and described in the annotation guidelines the label classes wall, door, window, floor, ceiling, pipe and pipe fitting will be presented here to describe the novelty of the annotation guidelines. Note, that per label specific entity descriptions, annotation rules and examples are given. In the following the full class definition and annotation guideline for walls will be given, door, window, floor, ceiling, pipe and pipe fitting will only be introduced briefly. Details can be found in the repository `https://gitlab.rhrk.uni-kl.de/kaufmann/humantech-data-annotation`.

To illustrate the descriptions screenshots of annotated data are presented. However, the given examples might not cover the exact same view and perspective since point clouds and images are different representations of the same

Figure 1: Examples of what invalid data can look like in images or point clouds: (left) Overexposure makes it impossible to distinguish different elements in an image and thus such areas are considered as invalid data. (right) Some points can be considered as noise and clutter in a point cloud and thus do not belong to any class and can be considered invalid data.

object and in most cases, point cloud and image annotation require specific rules according to the characteristics of the data and the annotation process. A particular annotation problem in image annotation might not exist for point cloud annotation and vice versa.

Walls

Walls generally consist of two parallel surfaces that are vertical by definition as described in the IFC entity definition: *"The wall represents a vertical construction that bounds or subdivides spaces. Walls are usually vertical, or nearly vertical, planar elements, often designed to bear structural loads. A wall is however not required to be load bearing."*

No additional description is required for walls, and examples of such annotations can be seen in Fig. 2 (left) and Fig. 2 (right) for images and point clouds respectively. Specific annotation rules for walls are defined:

- Preserve details on the edges and select them with utmost precision.
- Walls may be made of glass, and larger glass areas should be considered as walls.
- Typically, walls are supported by a floor or slab, and their vertical limit is the ceiling or bottom slab surface.
- Walls usually connect the floor and ceiling, except for openings and small edges.
- If a vertical member is connected to the floor but lacks a horizontal top surface below the ceiling, it is still annotated as a wall.

As mentioned, walls can also be made of glass or glass panes, which can be used for interior walls to separate rooms and spaces in a translucent way. In such cases, it is important to specify the difference between glass walls and doors. According to the proposed guideline, a door may include an area of fixed glass panes. If the area of the fixed glass panes or other panels around the door leaf does not exceed 1.5 times the door leaf area, it will be annotated as a door. In other cases where the fixed glass panes are larger than this threshold, they must be annotated as walls.

Skirting boards are part of the wall class. Unless there is a dedicated class for the objects, all objects attached to the wall such as pictures, light switches, etc. are annotated as wall.

Doors and Windows

Doors and **windows** can be distinguished by their nature given in the *IfcDoor* entity definition: *"The door is a building element that is predominately used to provide controlled access for people and goods. It includes constructions with hinged, pivoted, sliding, and additionally revolving and folding operations. A door consists of a lining and one or several panels,"* and windows are primarily used *"to provide natural light and fresh air,"* and may be horizontal or vertical such as skylights. While both door and window may contain a lining there are cases where only the glass pane will be fitted directly to the wall with-

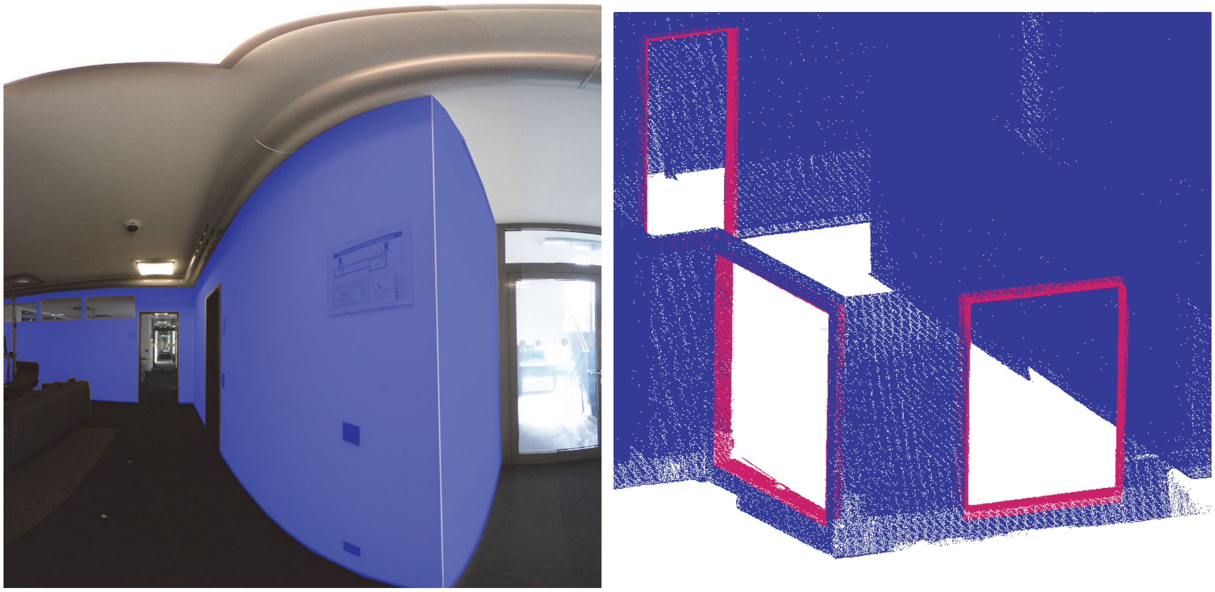

Figure 2: Examples of wall labels in images and point clouds: (left) Walls are labeled in blue, the glass panes around the door are not labeled as walls since they do not comprise more than 1.5 *times the door area. (right) Walls are labeled in blue, windows and doors are visible as openings in the wall.*

out a visible lining. In these cases only the panel would be annotated as window.

Technically, it is possible to annotate openings in images, although the mask may obscure objects visible through the opening. However, annotating openings in point clouds is not possible since there are only points on substantial objects and not on voids. Hence, void objects or classes that are not represented by data points cannot be annotated. Therefore, space or opening objects are not included in the proposed guidelines and are excluded from annotation, although these classes may be relevant for detecting falling hazards and other workflows. In later processes, such as scan-to-BIM workflows, openings must be identified based on the objects in which the openings are located, such as walls and slabs. In cases where a glass pane has no visible lining, it cannot be annotated in point clouds since glass panes are only present as artifacts, if at all.

Although glass panes are not visible in point clouds, they are visible in images. Since they are translucent, objects behind the glass pane are also visible, but may have slightly different textures due to minimal reflections, even from clear glass. Therefore, another notable feature of the guidelines is the labeling of transparent objects, such as glass doors or windows. In contrast to the approach by Armeni et al. (2017), which only labels the frame of a window or door while ignoring the transparent space, this guideline recommends annotating the glass area with a *transparency* flag as seen in Fig. 3.

Glass doors or windows can still be partially visible in Time-of-Flight (ToF) depth data or have different markings

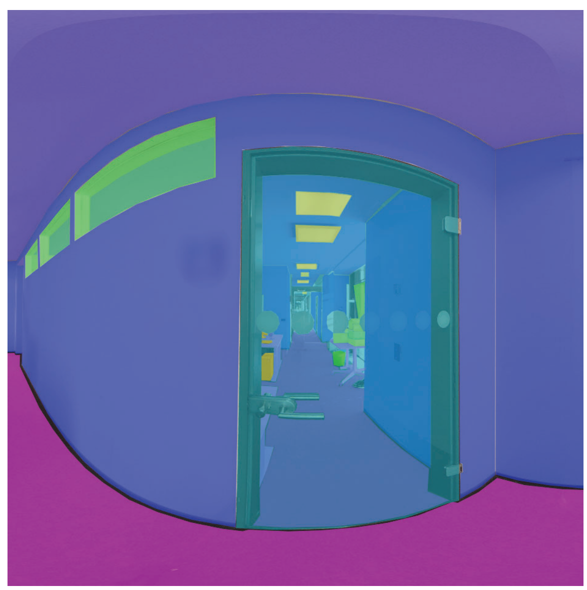

Figure 3: The transparent glass door as well as the objects behind are annotated.

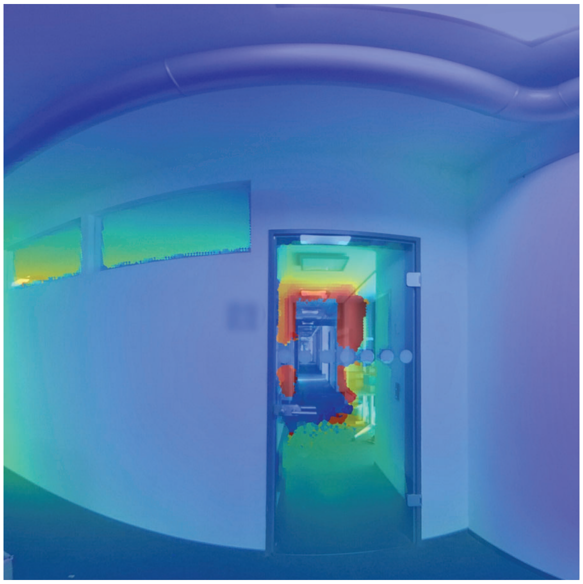

Figure 4: ToF depth overlaid on top of an image show that glass doors can be partially seen in the depth data.

on them, as shown in Fig. 4. The figure also shows that the objects behind glass are not only seen in the images due to transparency, but also their depth can be detected, partially or fully, making it necessary to label them as well with *occlusion* flag. The final recommendation is then to annotate both the transparent object as well as the objects behind it, ignoring any reflections, as can be seen in Fig. 3.

Floor and Ceiling

The examples presented above are strictly related to the IFC ontology as there are equivalent classes in the IFC standard. For many other classes such as furniture and construction equipment, material and even some building parts only generic IFC classes exist. As an example, the classes floor and ceiling will be explained in the following, many other examples can be found on the repository.

For both **floor** and **ceiling** the IFC class *IfcCovering* applies that is defined as *"an element covering other elements while being fully dependent on the element covered …"* (Buildingsmart, 2020). As this is not sufficient, additional descriptions are required to define the entity as follows:

- Floors are the bottom horizontal enclosing element of spaces.
- Ceilings are the top horizontal enclosing element of spaces. Ceilings include suspended ceilings and coverings of structural elements.

Both definitions indicate that the floor and ceiling cover the structural elements or installations, as per the *IfcCovering*, but also define that any layer other than the slab enclosing a space horizontally should be annotated as either floor or ceiling. Typically, both the floor and ceiling are horizontal, and inclined surfaces should be annotated with a different class, such as ramp or roof, if applicable. Only when the structural elements of the floor/ceiling construction are visible in the data should they be annotated with their respective classes, such as slab and/or beam. However, in point clouds, it can be difficult to distinguish these objects if there are no obvious visual features, such as rough concrete surface and texture. In such cases, additional documentation may provide more information about the state of the building, such as if it is still under construction.

Pipe Networks

Pipe networks consist of pipe segments and fittings. Hence, three classes are introduced to annotate respective data: Pipe segment vertical, Pipe segment horizontal and pipe fitting. Vertical and horizontal pipe segments are distinguished to facilitate a better segmentation and easier processing in scan-to-BIM pipelines. Pipe segments are used to *"typically join two sections of a piping network"* according to the IFC entity definition of the *IfcPipeSegment* class whereas an *IfcPipeFitting* connects individual pipe segments (Buildingsmart, 2020).

There is a specific challenge in annotating pipe fittings especially in point clouds. Depending on the point density, the separation line between the pipe fitting and the pipe segment might not be visible. In such cases, the pipe fitting ends where the axis is in the same direction of the pipe segment axis again i.e., when the curve of the pipe fitting ends. Due to the higher resolution of image pixels, it is likely that the separation line is visible in RGB data annotation. If not, the same rule applies. If the separation line cannot be identified, it should be annotated perpendicular to the pipe segment attached to the pipe fitting. Fig. 5 shows how pipes and pipe fittings can be annotated in both images and point clouds.

Data annotation

The task of data annotation is a labor intensive one that is done manually by multiple annotators. Ensuring uniformity in the labels is very important for the quality of the datasets produced. Therefore, the guideline not only serves as a basis for annotating data for the construction domain, but also guarantees the same quality of labels among different annotators.

RGB-D annotation

For RGB and RGB-D data annotation, tools such as COCO Annotator (Brooks, 2019) exist that allow for very fluent data annotation. Annotated data can be exported into quasi-standard dataset formats and directly used with these datasets in machine learning training pipelines. The annotations on the image data are polygons where each polygon is assigned one of the classes and other flags as needed

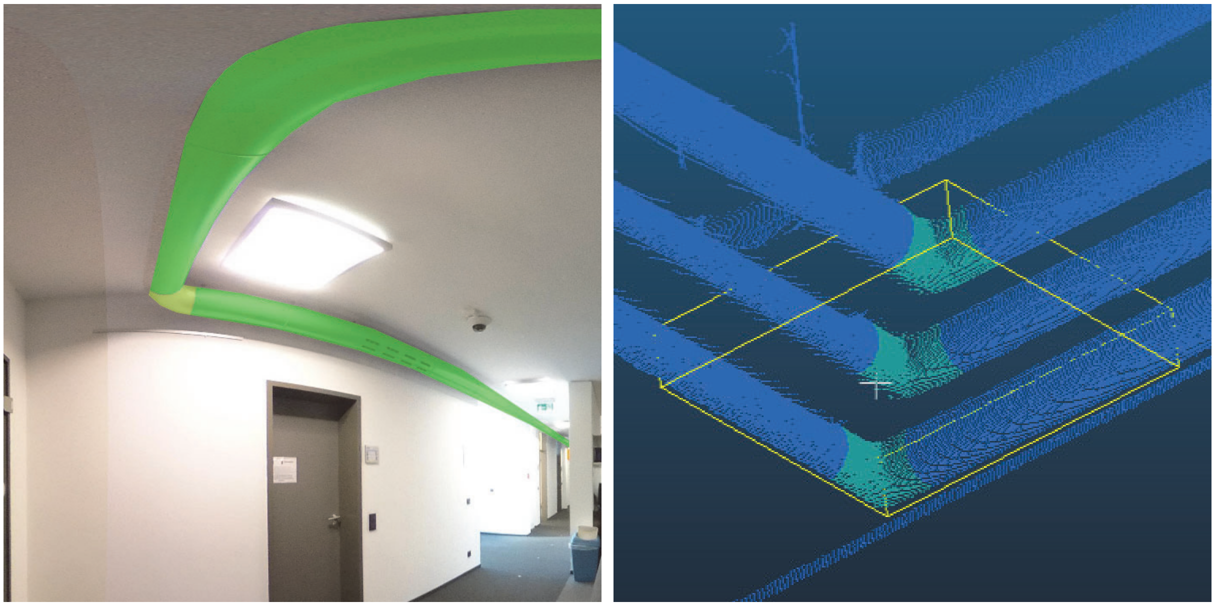

Figure 5: Examples of horizontal pipes and pipe fittings in point clouds: (left) Horizontal pipes and pipe fitting labels in image. (right) Pipes and pipe fittings are labeled in different colors in the point cloud.

(e.g. transparency in the case glass or occlusion for objects behind glass). The annotations are saved as json files.

Point cloud annotation

Semantic annotation of point clouds is a fully manual process. In the context of this work, CloudCompare (Girardeau-Montaut, 2020) is used. CloudCompare provides functions for the manual segmentation of point clouds. The labels are stored in a scalar field, i.e., an integer assigned to the respective point coordinate and color. The annotations can be visualized using color maps. For file transfer towards pre-processing and training pipelines, the PCD file format of the Point Cloud Library (Rusu and Cousins, 2011) is used as it can store points, coordinates, classification, and other values, combined with efficient compression using relatively small storage compared to plain text data formats.

Discussion and Future Work

While the guidelines introduced here will be used to annotate data, they will be amended and constantly improved. The more data that is annotated, the more specific conflicts will appear that can be solved by adding rules. The goal is to deliver a multimodal RGB-D and LiDAR dataset containing structural and non-structural building components and visual installations. If possible the dataset will be covered from empty buildings or buildings under construction to avoid occlusions from interior objects such as furniture. The raw data was captured with RGB-D and LiDAR sensors under equal environment and lighting conditions. The data will be annotated using the proposed annotation guidelines and will facilitate the comparison of image-based and point-cloud-based segmentation and BIM generation methods as well as general data comparison.

As introduced in the related work section, other datasets exist already focusing on different kinds of buildings and objects. To use these datasets fluently in training pipelines, label mapping tables could be elaborated. This would facilitate the use of much more data. Besides label mapping, a dataset focusing on structural components in buildings, ideally obtained from empty buildings, could provide the basis for semantic reconstruction of such elements. This would serve a wide range of use cases, such as obtaining BIMs of existing structures for re-design or acquiring material quantities in existing structures to assess opportunities for recycling and reuse.

So far, this work has focused on completed buildings and their data annotations. However, in an ongoing construction project, objects such as partially completed walls and ceilings can be frequently encountered. Therefore, it would be advantageous to integrate such structural components for spatio-temporal aspects as well. This work can be extended to include data modalities like LiDAR.

Conclusions

In this work, a unified guideline for RGB-D and point cloud data annotation based on construction ontologies is presented and published. This will serve as the foundation for labeling large amounts of data and, as a result, supervised training of deep neural networks for scan-to-BIM automation and the generation of digital twins. It will also improve the compatibility between different datasets that follow the same guidelines and thus allow for training on data from multiple datasets.

Acknowledgments

This research is part of the HumanTech project (humantech-horizon.eu) and was partially funded by the European Union under the Grant Agreement 101058236.

References

Armeni, I., Sax, S., Zamir, A. R., and Savarese, S. (2017). Joint 2d-3d-semantic data for indoor scene understanding. arXiv preprint arXiv:1702.01105.

Armeni, I., Sener, O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., and Savarese, S. (2016). 3d semantic parsing of large-scale indoor spaces. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1534–1543.

Brooks, J. (2019). COCO Annotator. https://github.com/jsbroks/coco-annotator/.

Buildingsmart (2020). Industry Foundation Classes 4.0.2.1. https://standards.buildingsmart.org/IFC/RELEASE/IFC4/ADD2_TC1/HTML/.

Chang, A., Dai, A., Funkhouser, T., Halber, M., Niessner, M., Savva, M., Song, S., Zeng, A., and Zhang, Y. (2017). Matterport3d: Learning from rgb-d data in indoor environments. arXiv preprint arXiv:1709.06158.

Construction Specification Institute (2023). OmniClass. https://www.csiresources.org/standards/omniclass.

Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T., and Nießner, M. (2017). Scannet: Richly-annotated 3d reconstructions of indoor scenes. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5828–5839.

Geng, J. (2011). Structured-light 3d surface imaging: a tutorial. Advances in Optics and Photonics, 3(2):128–160.

Girardeau-Montaut, D. (2020). CloudCompare: 3D point cloud and mesh processing software.

Gupta, S., Arbelaez, P., and Malik, J. (2013). Perceptual organization and recognition of indoor scenes from rgb-d images. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 564–571.

Gupta, S., Girshick, R., Arbeláez, P., and Malik, J. (2014). Learning rich features from rgb-d images for object detection and segmentation. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13, pages 345–360. Springer.

Kaufmann, F., Glock, C., and Tschickardt, T. (2022). Scalebim: Introducing a scalable modular framework to transfer point clouds into semantically rich building information models. In EC3 Conference 2022, volume 3, pages 0–0. University of Turin.

Lin, D., Fidler, S., and Urtasun, R. (2013). Holistic scene understanding for 3d object detection with rgbd cameras. In Proceedings of the IEEE international conference on computer vision, pages 1417–1424.

National Building Specification (2022). Uniclass - Unified Construction Classification. https://www.thenbs.com/our-tools/uniclass.

Ren, X., Bo, L., and Fox, D. (2012). Rgb-(d) scene labeling: Features and algorithms. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 2759–2766. IEEE.

Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2005). Labelme: a database and web-based tool for image. Int. J. of Computer Vision, 77(1).

Rusu, R. B. and Cousins, S. (2011). 3d is here: Point cloud library (pcl). In 2011 IEEE international conference on robotics and automation, pages 1–4. IEEE.

Silberman, N., Hoiem, D., Kohli, P., and Fergus, R. (2012). Indoor segmentation and support inference from rgbd images. ECCV (5), 7576:746–760.

Song, S., Lichtenberg, S. P., and Xiao, J. (2015). Sun rgb-d: A rgb-d scene understanding benchmark suite. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 567–576.

Zheng, J., Zhang, J., Li, J., Tang, R., Gao, S., and Zhou, Z. (2020). Structured3d: A large photo-realistic dataset for structured 3d modeling. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16, pages 519–535. Springer.