Review of Window and Door Type Detection Approaches

Sam De Geyter[1,2], Maarten Bassier[1], Heinder De Winter and Maarten Vergauwen[1]

[1] Dept. of Civil Engineering – Geomatics, KU Leuven – Faculty of Engineering Technology, Ghent, Belgium
(sam.degeyter, maarten.bassier, heinder.dewinter, maarten.vergauwen)@kuleuven.be
[2] MEET HET BV, Mariakerke, Belgium

Technical Commission II

KEY WORDS: Scan-to-BIM, BIM Parameterization, Window and Door Modelling, Remote sensing data

ABSTRACT:

The use of as-built Building Information Models (BIM) has become increasingly commonplace. This process of creating a BIM model from point cloud data, also referred to as Scan-to-BIM, is a mostly manual task. Due to the large amount of manual work, the entire Scan-to-BIM process is time-consuming and error prone. Current research focuses on the automation of the Scan-to-BIM pipeline by applying state-of-the-art techniques on its consecutive steps including the data acquisition, data processing, data interpretation and modelling. By automating the matching and modelling of window and door objects, a considerable amount of time can be saved in the Scan-to-BIM process. This is so because each window and door instance needs to be examined by the modeller and must be adapted to the actual on-site situation. Large object libraries containing predefined window and door objects exists but the matching to the best-fit predefined object remains time consuming. The aim of this research is to examine the possibilities to speed up the modelling of window and door objects. First, a literature review discussing existing methods for window and door detection and matching is presented. Second, the acquired data is examined to explore the capabilities of capturing window and door information for different remote sensing devices. Followed by tests of some commonplace features in the use for window and door occurrence matching and clustering.

1. INTRODUCTION

The use of Building Information Modelling (BIM) is rapidly gaining terrain within the Architecture, Engineering, Construction (AEC) and Facility Management (FM) industry. For existing buildings, as-built BIM models are needed. These models are mostly created using detailed 3D measurements of the building. The Scan-to-BIM process is mostly manual which makes it time-consuming and error prone, and thus inefficient. The BIM reconstruction of window and door objects and secondary building components in general is more complicated than the geometry extraction of primary building components such as walls and floors. The main cause for this extra level of difficulty is the often complex geometric shape of these objects where primary building components mostly exists of large surfaces, this is not the case for window and door objects. These objects mostly need a more detailed geometry including multiple sub components such as frames, handles etc. Additionally, the remote sensing data used for the modelling of these objects typically lack information about these specific elements. For instance, occlusions caused by incomplete scanning or other objects blocking the line of sight to crucial parts are a major problem

In this work, the process of identifying distinct types and clustering similar occurrences of window and door objects within a project is examined. Currently, modellers examine both the point cloud and image data to identify all distinct types of windows and doors within the project. Doing this for every window and door object within a project is a time-consuming task. This process can be optimized by clustering all occurrences of one type and suggest a matching element from pre-existing libraries, containing huge amounts of BIM objects. Especially because a project only contains a limited number of different window and door object types. For example, in the university building in the city of Ghent only six types of windows represent over 51% of all windows in the entire BIM model. The same for doors were also six types represent 52% of all doors as shown in Figure 1. Tools to identify and group these distinct types would

already significantly speed up the interpretation of the remote sensing data. Where an estimated 60% of the modelling time is spend searching for accurate already existing models in object libraries (Li et al., 2020). For the purpose of identifying and clustering similar occurrences, a distinct descriptor or method to compare two occurrences for those objects is needed.

The amount of information needed by a modeller to create a BIM object is dependent on the project deliverable. This work only focuses on geometric data which can be captured using remote sensing techniques. The extraction of non-geometric parameters lies beyond the scope of this research as these cannot be determined from remote sensing data. According to the Level of Development standards (Bedrick et al., 2020) the amount of information in a model can be expressed using LODs. Where the LOD200 standard for windows and doors only requires placeholders representing the approximate location and size of the window or door instance. Most projects demand window and door objects to be modelled up to the LOD300 specification. This requires the exact geometric representation of the window and door object.
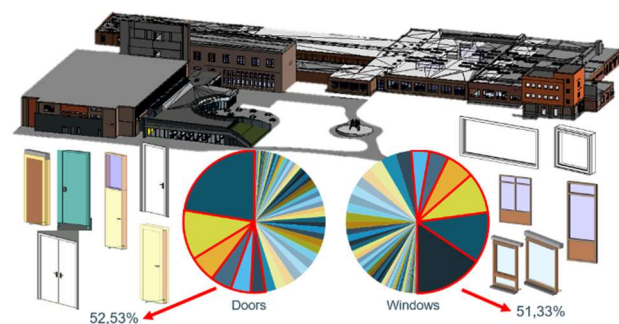

Figure 1: Distribution of window and door types on the University building in Ghent. Showing six types of each represent more as half of all window and door objects within the project.

In practice, most of the window and door types can be represented by an already existing model and some unique types must be modelled from scratch. The time needed to match or model a window or door occurrence will strongly depend on the uniqueness and complexities of the window or door type. To this end, this work examines the possibilities to automate the matching of different occurrences of window and door types together. By grouping all occurrences of each window or door type a targeted search for each window or door type represented in the project can be conducted instead of a manual search.

The presented work is structured as follows. Section 2 containing an overview of recent literature concerning the needed steps for window and door reconstruction for Scan-to-BIM. Followed by a section discussing the difficulties concerning the capturing, interpretation and shortcomings of remote sensing data for window and door extraction. Section 4 and 5 discuss the methodology of some conducted experiments and their results. The following sections present a method for clustering of different occurrences of similar objects in image data. Finally section 7 concludes this initial work and outlines the future work towards an (semi-)automated window and door BIM reconstruction method.

2. RELATED WORK

The related work discussed in this section is subdivided in two parts. The subsection discusses the entire pipeline from window and door detection to the placement of window and door objects from existing libraries. The second part of this related work focusses more on the extraction of features from 3D shape representations and how features of different elements are compared.

2.1 Window and door detection and modelling

The detection of window and door objects in remote sensing data is widely researched. (Neuhausen et al., 2016) gives an overview about the different approaches targeting image-based window detection focussed on facades. They distinguish three main detection approaches. (1) Grammar-based approaches using a set of predefined rules and assumptions. (2) Pattern recognition approaches based on the mostly grid-like structure of windows parallel to the facade contour or sets of parallel vertical and horizontal lines. (3) Machine learning techniques targeted to detect window and door objects in images and even point clouds. The first approach is challenging to generalize since the rules must be adapted to each type of building to yield decent results. When the rules are too generic, the algorithm will only split the façade in rectangles where too complex rules narrow the applicability. Pattern recognition approaches usually also rely on some assumptions for example the differences in pixel intensity (Musialski et al., 2013) or assumptions made regarding the shape and alignment of windows on facades (Van Gool et al., 2007).

The use of machine learning techniques on images for per pixel predictions is discussed in (Long et al., 2017). (Schmitz and Mayer, 2016) use an end-to-end learned convolutional Neural Network (CNN) for image segmentation on facades. This fully convolutional approach was extended with optional man-made rules concerning the typical symmetry found in structures reporting segmentation accuracies for windows and doors of 93.04% and 90.95% respectively on the S3DIS dataset (Liu et al., 2020). State-of-the-art deep learning methods tackle the task of three dimensional semantic segmentation on point cloud data (Qi et al., 2017a, 2017b). One of the more recent works by (Hu

et al., 2020) using random sampling for large scale point clouds. Compensating the loss of information caused by the random sampling by introducing a local feature aggregation module and enabling the preservation of important geometric details. Their network achieves mIoU results on windows and doors in the S3DIS dataset of 64.6% and 69.4%.

Other networks leverage the use of DL on both types of inputs. (Robert et al., 2022) introduce DeepViewAgg where the images and the point cloud data are processed by dedicated DL networks in an end-to-end fashion. Merging features of images taken from different positions based on the 3D points viewing conditions, this method achieves mIoU results for windows and doors of 71.9% and 78.9% on the S3DIS dataset. Other works combining point cloud data and RGB images focus on door detection and compute their opening angles (Quintana et al., 2018). First, measures are discussed to ensure the quality of the colour information such as using flash, removing specular highlights and merging colour information from different views. The proposed technique starts by detecting the wall surface in the RGB-D space by clustering pixels which are coherent in both RGB and Depth. By imposing thresholds on all four components and assuming the wall area to contain the largest number of pixels on the side and top borders of the image, the wall surface is extracted. The doors are extracted by searching discontinuities in this 4D RGB-D space. This method shows promising results on the determining of the door location, size for closed and opened doors even in occluded conditions.

(Adán et al., 2018) takes this idea further, searching for even smaller components such as power outlets, switches, fire safety equipment etc. in the captured data. Their method requires the wall model as additional input data and a pre-existing object library containing colour and depth image models of the objects present in the scene. The input data is then split into its RGB space and its depth and processed separately. The depth information is used to generate regions of interest using discontinuities and a Canny filter to detect the edges. The depth information contained in the region of interest is then matched to a depth signature of the library using an image cross-correlation algorithm to asses the match. The RGB data is processed in a similar way detecting discontinuities in the colour domain. For the image based regions of interest a distance-based classifier using global descriptors invariant to scale and rotation is used for matching against the library objects. The authors do not use local features such as (SIFT, SURF, etc.) because these do not produce good results on their orthoimages due to an insufficient resolution. To evaluate the results of both methods, the results for every region of interest are used to compute a Recognition
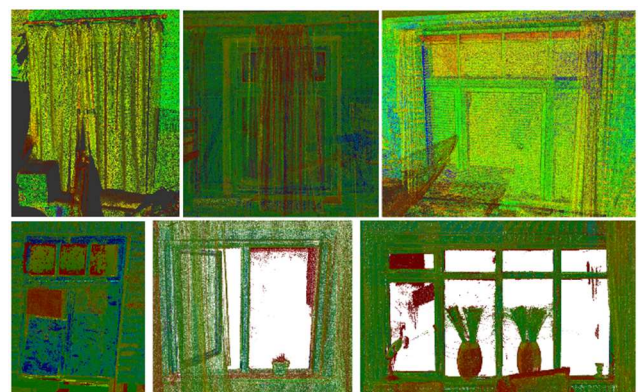

Figure 2: Point Clouds of windows showing typical obstructions preventing information extraction for window reconstruction

Coherence Matrix to reach a final match for which the position is computed (Adán et al., 2018).

Before objects can be matched with modelled objects from a library, BIM object models must be stored in such a library. Nowadays most companies using BIM have such libraries available. Additionally, there are the online object model libraries i.e. BIMobject[1], C3A library[2] and Bimsmitch market[3] that already contain thousands of AEC object models. Each of these libraries maintain their own categorization, mostly dependent on user keywords, which lacks standardization. As a result, a large number of objects is rarely used again (Iyer et al., 2005). (Abdirad and Mathur, 2021) reports that searching for the correct BIM object in libraries approximately takes 2-4% of annual working hours. To combat this issue, BIM content management and delivery (BCMD) techniques are investigated to standardize naming, quering and organisation. Additionally, ML quering is proposed to retrieve BIM object models more effectively during design (Abdirad and Mathur, 2021). This is supported by (Li et al., 2020), who discuss how to more efficiently build 3D models from scratch and enable the reuse of already existing object models.

2.2 Feature extraction and matching

In the Scan-to-BIM context, the main sources of information for type detection are point clouds and RGD imagery of the captured object. To enable matching with an already existing library object, a distinct description of the objects' geometric representation and matching method between captured and library object is needed. Many works presented in literature rely on the use of handcrafted features and heuristics to match similar object representations (Krishna et al., 2021). This section will discus the two main feature types being Geometry-based and view-based features (Bickel et al., 2022).

Geometry-based methods use the geometric information to determine geometry descriptor or feature of the 3D shape. Typical global 3D features include Shape histogram features (Ankerst et al., 1999) or Shape Distribution features (Osada et al., 2001), where random sampling is used to create a continuous probability distribution which is used as a signature for the 3D shape. Local 3D geometric features are Regional point descriptors (Frome et al., 2004), Extended Gaussian Images, Conformal Factors, Spherical Harmonics and Poisson histogram descriptors (Li et al., 2015). A mix of both a descriptor types is also proposed. For instance, (Sipiran et al., 2013) presents a mesh partitioning scheme where a global mesh descriptor is used in combination with local mesh partition descriptors. This approach helps to retain the local geometric information which is often lost by using global features.

View-based 3D model retrieval. Where this method can benefit from the advanced existing image processing techniques (Liu, 2012). This technique is based on rendered images of the 3D geometry where afterwards the similarity between the rendered image and library object is computed (Bickel et al., 2022).

Recent advances in the field of ML, Deep feature representations of point cloud data are created using various ML techniques. A feature vector of the point cloud can be created by for example extracting a per point hierarchical feature and a Self-Organizing Map (SOM) to capture the spatial distribution of the point cloud (Li et al., 2018). A extensive overview of different DL methods for 3D geometry representations such as multi-view images, voxels, point clouds, meshes etc. is given in (Xiao et al., 2020). They argue that multi-view image representations preform better than other 3D representations due to the learning methods and the success of 2D DL networks. This is confirmed by (Bickel et al., 2022) stating the extensive knowledge of pretrained DL networks from the field of computer vision can be used enabling the creation of a shape descriptor through feature extraction without the need of labels. An example, where different views are separately processed and then joined using view-pooling layers is presented in (Su et al., 2015). Other methods using point cloud data as input come back to the networks mentioned in section 2.1.

An example of object retrieval approach in the AEC industry is BIMSeek++ (Li et al., 2020). Using the BIMSEEK environment similar objects from an existing BIM library are extracted using an attribute input containing the 3D shape (IFC). For matching between the objects the Tversky similarity is used to retrieve similar already existing models containing domain specific information. Another approach using the geometric representation of an object to search for similar objects in a library can be found in the field of mechanical engineering (Bickel et al., 2022). They project both the observed point cloud and the geometry of the objects in the library onto a sphere subdivided in pixels and counting the number of projections per pixel. By doing this for different rotations of the objects and unfolding these spheres to matrices feature vectors are extracted using DL on the different matrices for an object. The matching is performed by conducting a k-nearest-neighbour search for these feature vectors to determine the best matching object.

3. WINDOW AND DOOR DATA DISCUSSION

Besides their complex geometries window and door objects inherently suffer from data shortcomings of the remote sensing data as shown in Figure 2.

Typically glass surfaces are places in the point cloud data containing less points. This lack of data can be seen as data on itself but makes it almost impossible to determine the thickness of the glass plate itself. The window frame contains important information for the window reconstruction. This frame which is typically manufactured of materials that can be captured by remote sensing techniques suffers from the limitations of these techniques itself. For the frame reconstruction the position and settings which are used during the data capturing can be of large influence. The position of the scanner will have a direct impact on the visibility of the frames which are mostly occluded by the wall it is contained in. Also the distance to and the resolution of the scanner will have a direct impact on the number of points captured on the frame itself. This scanner dependent limitations can be solved by using mobile scanners which will typically obtain a higher coverage of such areas where Terrestrial Laser scanners will need a large number of additional setup locations to capture the same amount of data (De Geyter et al., 2022).

Besides the occlusions and data gaps caused by the sensor or the structure itself additional clutter objects can cause data gaps. These gaps are mostly situated on the window frame, so hiding important information for the window reconstruction. For

---

[1] https://www.bimobject.com/nl-be
[2] https://www.c3a.be/c3a-bim-families
[3] https://market.bimsmith.com

example window decorations such as curtains typically cover the sides of the window where essential geometric information of the frame is located.

Some of these above mentioned problems are solved by using RGB imagery. These images can be taken by the laser scanner from the same positions but in that case will suffer from the same data occlusions as the scanning data. Inherently, the usability of these image data are also highly dependent on lighting conditions. The cameras provided by most laser scanning systems do not have the ability to influence the lightning conditions of the scene and therefore will not be able to generate much extra information besides filling the gaps caused by low point densities. Additionally the resolution of the images taken by the laser scanner is typically low, this strengthens the motivation to capture more images with other camera configurations. Capturing additional images from other position can provide additional data for the reconstruction. Also, when another cameras as the camera of the laser scanner is used the lightning conditions can mostly be adapted resulting in better colour and texture information as well as better visibility by reducing shadows.

Additionally, the state of the targeted object is rarely fixed. During capturing it isn't uncommon that door or window objects change from an opened state to a closed state which makes it more difficult to detect similar objects both in images as point cloud data. This problem remains when matching similar occurrences or when matching against library objects.

From these limitations it can be concluded that both types of input data are needed to extract the needed geometric information. A possible solution for the lack of information is the use of the occurrence clustering explained in section 6. By clustering different occurrences of the same window or door types and merging these matched occurrences together it should be possible to restore the entire type representation and extract all needed data.

4. REVIEW OF TYPE DETECTION METHODS

In this section, three promising matching methodologies are adapted to detect the proper object type from a pre-existing BIM object library given a set of classified point cloud and image observations. Concretely, we evaluate the matching of (1) geometric density signatures of point cloud data to the shape of BIM elements, (2) geometric registration features of point cloud data to the shape of BIM elements and (3) image correlation features to orthophoto's of BIM elements.

4.1 Geometric features

For the geometric processing, a set of point clouds $Q = \{Q_1, Q_2, …, Q_n\}$ are sampled on the potential BIM element types with a fixed spatial resolution. The same spatial resolution is applied to the classified point cloud P that originates from any RGBD or Lidar sensor. Note that Q is conditioned to only contain points that are theoretically visible to the sensor (Fig.3). This includes a filtering of the observable materials i.e. the glass panels and the removal of points on the object's interior i.e. in joints or embedded in the wall. Additionally, a small portion of the surrounding wall is sampled and added to each $Q_n \in Q$ to better resemble the classified observations.

4.1.1 Density signatures

The first method leverages point density histograms to determine the best fit $Q_n \in Q$ for a given P. To this end, three histograms are extracted from every Q_n and P along their respective cardinal axes (Figure 5). The first axis is defined parallel to the wall, the second axis is along the Z-axes and the third axis orthogonal to wall face. The histograms h_p and h_q are then obtained by respectively projecting P and Q_n on these axes. The cosine distance between the vectors of each histogram is used as the metric to determine the best fit match between P and $Q_n \in Q$ (Eq.1).

$$D = 1 - min_T \sum_i T h_p(i) h_q(i) \qquad (1)$$

Where i is the same number of bins for each histogram. To align the histograms and to limit the impact of occlusions, a best fit transformation T is computed between both histograms that minimizes the cosine distance.

4.1.2 Registration features

The second method leverages both local and global geometric registration methods. First, a conditional Iterative Closest Point algorithm (ICP) is considered. To this end, the center point of P is positioned at Q and P is rotated around the Z-axis so that the dominant axis of their bounding boxes are aligned in the XY-plane. Next, the best fit rigid body transformation is determined between both point clouds (Eq.2).

$$RMSE = min_T \sum_P \rho(\|p - Tq\|) \qquad (2)$$

Where the RMSE and percentage of inliers determined by ρ are used to determine the best fit match between P and Q. Second, a global registration is considered based on geometric features. To this end, a set of FPFH features is extracted from both P and Q as implemented by (Zhou et al., 2016). A RANSAC variant is then applied to retrieve the best fit transformation between P and Q. Analogue to the ICP, the model fitness (% feature inliers) and RMSE between inlier points are used to determine the best fit BIM object type. Note that no refinement step is performed to limit the influence by occlusions and noise.

4.2 Image correlation features

The third method computes image features for the matching. For the image processing, virtual images $I = \{I_1, I_2, …, I_n\}$ are generated from the potential BIM element types and the occurrences in the found occurrences in the point cloud data.
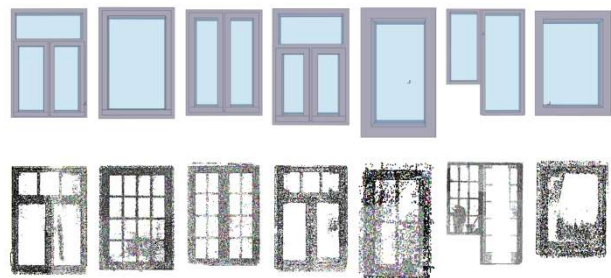

Figure 3: Selection of used window BIM elements and point clouds

	Local registration						Global registration					
	Registration RMSE (m)			Registration fitness (% inliers)			Registration RMSE (m)			Registration fitness (% inliers)		
	Correct model	Lowest value	Highest value	Correct model	Lowest value	Highest value	Correct model	Lowest value	Highest value	Correct model	Lowest value	Highest value
Cloud 1	0.0154	0.0154	0.2654	100	65.83	100	0.0123	0.0095	0.0123	23.52	15.51	38.03
Cloud 2	0.017	0.015	0.221	100	100	100	0.011	0.0096	0.0111	65.71	35.68	77.46
Cloud 3	0.0132	0.0132	0.326	100	87.67	100	0.011	0.0103	0.011	51.52	13.25	51.52
Cloud 4	0.011	0.011	0.211	100	100	100	0.0108	0.0093	0.0112	37.53	13.06	70.79
Cloud 5	0.0179	0.0136	0.1991	100	100	100	0.0107	0.0099	0.0116	64.03	14.46	68.5
Cloud 6	0.0173	0.0109	0.184	100	100	100	0.0104	0.0103	0.0111	48.23	25.72	76.43
Cloud 7	0.0117	0.0117	0.0246	100	100	100	0.0106	0.0103	0.0113	78.38	35.07	78.38
Cloud 8	0.0106	0.0095	0.0283	100	100	100	0.0103	0.0091	0.0117	85.43	46.3	94.19
Cloud 9	0.0105	0.0105	0.0198	100	100	100	0.0114	0.0099	0.0114	64.78	25.92	84.09
Cloud 10	0.0169	0.0169	0.3103	100	46.65	100	0.0114	0.0098	0.12	13.05	7.3	40.93

Table 1: Summarized registration results, reporting the value of the correct match and the highest and lowest value for each parameter and both registration approaches

The matching itself is based on state-of-the-art deep learning features. Concretely, we extract 1280 image correlation features using a freely available EfficientNet (Tan and Le, 2019) adaptation that is trained on ImageNet. Analogue to the geometric processing, We compute the cosine distance between the feature vectors of I_p and every $I_n \in I$ as a metric for the image matching. In addition to the raw images, we also conduct the same test for lines extracted from the initial imagery. To this end, we extract the Sobel image using X- and Y-gradients in the image.

5. EXPERIMENTS

Two datasets are used for the experiments. The first dataset consists of point clouds of windows in a residential building captured with a Terrestrial Laser Scanner (Figure 3). The second dataset consists of interior and exterior images of windows of the university building in Ghent captured by DSLR (Figure 4). For completeness, images taken from the point cloud data and BIM are also included. Following the data discussion in section 3, only the three most frequently occurring and automatable windows and doors are selected. Two types have a limited height of approximate 0.5 meters, one existing of 3 sub windows and one existing of only one almost square window. The third type consists of 6 sub windows configured in two rows of three each, the lower row is smaller as the upper row. Both datasets have all the frequently occurring shortcomings due to occlusions, clutter, etc. as discussed in section 3. For the BIM to point cloud virtual image matching only the first two types are used.

The results of the geometric registration are shown in **Fout! Verwijzingsbron niet gevonden.**. The spatial resolution of the unification and sampling of the point clouds is 0.010m. On average, RMSE and fitness values for the correct matches are respectively 0.014m and 100% for the local registration method. For the global registration method the results are 0.011m and 53.22%. For erroneous matches this is 0.097m and 95.94% for the local approach and 0.012m and 41.97% for the global registration. Only in 6/10 matches, the correct match has the lowest RMSE and highest fitness using the local approach. However, the fitness did not provide a unique solution. With the

Figure 4: Selection of used images. Containing RGB images of the real life objects, renders of the BIM model and Point cloud.
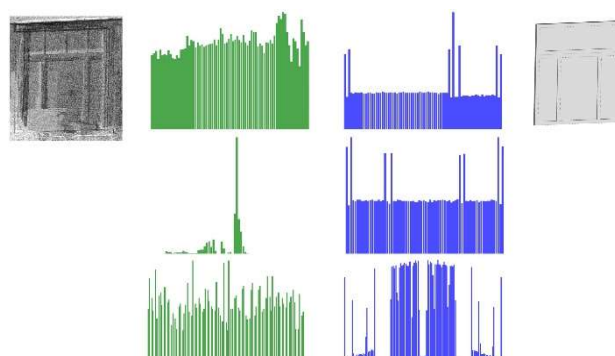

Figure 5: Histograms of the captured point cloud (green) and sampled point cloud of the BIM model (Blue) in each direction.

| Point cloud renders | Features | | | | | | Sobel images with extracted lines | | | | | |
| | Correct matches Similarity [m] | | | Wrong matches Similarity [m] | | | Correct matches Similarity [m] | | | Wrong matches Similarity [m] | | |
	Average	Lowest value	Highest value	Average	Lowest value	Highest value	Average	Lowest value	Highest value	Average	Lowest value	Highest value
Model 1	0.46	0.33	0.38	0.52	0.38	0.66	0.55	0.31	0.41	0.54	0.47	0.63
Model 2	0.31	0.29	0.31	0.56	0.45	0.72	0.50	0.45	0.52	0.43	0.40	0.50
Images												
Model 1	0.74	0.60	0.82	0.80	0.76	0.84	0.62	0.47	0.72	0.64	0.54	0.77
Model 2	0.74	0.70	0.75	0.74	0.59	0.80	0.52	0.46	0.55	0.57	0.46	0.69

Table 2: Summary of the similarity computations between rendered images of the BIM elements and the point clouds and the images.

global approach no matches where result of the lowest RMSE and the highest fitness. Nevertheless, 2/10 cases the highest fitness represented the correct match. A key issue is that smaller and less detailed window/door observations also resonate well with larger and more intricate BIM elements. The window configuration also has limited impact on the matching as there are too few geometric features on these details. Overall, it can be concluded that both ICP and feature-based geometric matching is crucially flawed for window/door type matching.

The results of the density histograms are shown in Figure 5. The spatial bins of the histograms are 0.010m. On average, the similarity values are 0.30 for the correct matches and 0.32 for the erroneous matches. However, only in 3/10 matches, the correct match has the lowest similarity. Additionally, on average there is only a 0.02 difference between the correct and erroneous matches. Even though the BIM elements have a widely different geometry. The key issues are (1) occlusions, which cause confusion in the best fit transformation and significantly lower the similarity, (2) interference of clutter that cause unexpected peaks and (3) a critical amount of noise on the glass panels and due to ghosting (4) the scale invariance.

The results of the image feature matching are shown in Table 2. All images were cropped to 224x224 pixels. For the feature matching between the rendered BIM and PCD, the cosine distance values are on average 0.38 for the correct matches and 0.54 for the erroneous matches. For the Sobel Images with the extracted lines, the distances are on average 0.52 for the correct matches and 0.52 for the erroneous matches. images, 9/10 matches where correct. However, with the Sobel images with extracted lines only one type was successfully matched resulting in 6/10 matches. When comparing RGB images to the BIM renders, using feature matching or Sobel images with line extraction, the average cosine distance values are respectively 0.74 and 0.57 for correct matches and 0.74 and 0.46 for erroneous matches. In both image vs BIM render cases, only 3 images were matched to the correct BIM. Using the feature method only one type was recognized, but two images where wrongfully classified. For the Sobel image with extracted lines, both types where recognized but high confusions between both types occur.

Overall, it is stated that image correlation methods significantly outperforms geometric matching methods. Regardless, all above approaches prove insufficiently viable even for this straightforward testcase. The key culprits are shortcomings in the data rather than the methods themselves. As documentation methods are unlikely/unfavourable to change, a solution must be found in the data pre-processing to better define the descriptor and matching task.

6. OCCURANCE CLUSTERING

We propose to improve the type detection by clustering the image and point cloud observations per type prior to the matching. The hypothesis is that while the above methods fail to reliably match an observation to an as-design library object, these methods will significantly perform better by matching observations to observations. There is less hinder from texture confusion, model abstractions, sensor positions and so on. This is also extremely relevant as over half of all window and door objects can be represented with only a few object types. Once a set is clustered, the matching of individual observations can be aggregated to assign majority vote to the entire cluster. Furthermore, once a type is selected, the separate occurrences can be used to optimize the parameter estimation incl. the width, height, depth, panel configuration and so on.

To test the hypothesis, we conduct an initial test with the image based matching. The same images used in the above experiments are cross-referenced with each other and their similarities evaluated as discussed in Section 4. To this end, a densely connected topology graph is constructed with $|I|$ the number of nodes. The edges are defined by the cosine distance between the feature vectors as discussed in Section 4. A connected component analysis is performed on the graph. Edges are considered valid if the cosine distance are within an empirically determined threshold t_c of 0.3 (Eq.3).

$$\sum_{i,j} f(I_i)f(I_j) \leq t_c \qquad (3)$$

Next, the matching type is selected by the majority vote within each cluster. For this initial testcase on 23 images of three window types, the results are promising. In total, 7 clusters are created of which 21/23 images are assigned to a cluster and 2 images are not assigned to any cluster. Five of these clusters select the correct type with all of the votes being correct. This is a significant improvement over the individual matching. The larger the cluster the larger the gain. In contrasts, clusters with a single image receive no benefit from the matching. Several improvements will be explored in future work i.e. evaluating similarities between clusters, reducing the combination complexity of the graphs, embedding geolocation between images and point clouds and so on.

7. DISCUSSION & FUTURE WORK

This paper presents a review of window and door object type detection for scan-to-BIM procedures. Three methods are evaluated to match close-range remote sensing data i.e. point cloud and RGB imagery to BIM object types in pre-existing libraries. Concretely, two geometric and one image-based matching method are tested including density histograms, registration features (ICP & FPFH) and DL image features. Additionally, a framework is outlined to cluster the observations prior to the type matching to improve the detection rate. The main contribution is the assessment of state-of-the-art descriptors to encode and reliably match remote sensing observations to objects in pre-existing libraries

The literature study presented in this work concludes that the use of image-based features for 3D shape matching is promising where the geometric methods struggle to efficiently describe the objects in sufficient detail. This is confirmed in the preliminary tests that show that both local as global registration parameters and point density histograms don't yield promising results. The use of rendered images of both the BIM and the point cloud on the other hand yield more promising results. Methods of RGB imagery matching against BIM renders do not perform as expected. A potential problem can be found in the lack of texture of in the BIM renders. This supports the conclusion of the presented literature study opting for image-based matching techniques.

The use of feature lines extracted from RGB images is explored to match similar windows or different occurrences of the same image type. The tests presented in this work use the cosine distance between the image feature line representations. From these tests, it can be concluded that the use of this method to cluster similar images is promising. By clustering multiple occurrences of the same window or door type, information gaps caused by occlusions can be filled in by merging data from these different occurrences.

Future work is needed to further explore the use of both 3D and 2D view-based matching techniques and features. Where view-based techniques look promising the continues increase of deep learning networks for 3D data needs to be watched. Where these networks increase their ability to create distinct both local and global features which can be used for matching.

ACKNOWLEDGEMENTS

REFERENCES

Abdirad, H., Mathur, P., 2021. Artificial intelligence for BIM content management and delivery: Case study of association rule mining for construction detailing. Adv. Eng. Informatics 50, 101414. https://doi.org/10.1016/j.aei.2021.101414

Adán, A., Quintana, B., Prieto, S.A., Bosché, F., 2018. Scan-to-BIM for 'secondary' building components. Adv. Eng. Informatics 37, 119–138. https://doi.org/10.1016/j.aei.2018.05.001

Ankerst, M., Kastenmüller, G., Kriegel, H.P., Seidl, T., 1999. 3D shape histograms for similarity search and classification in spatial databases. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 1651, 207–226. https://doi.org/10.1007/3-540-48482-5_14

Bedrick, J., Ikerd, W., Reinhardt, J., 2020. Level of Development (LOD) Specification Part I & Commentary For Building Information Models and Data. BIMForum 263–64.

Bickel, S., Schleich, B., Wartzack, S., 2022. A Novel Shape Retrieval Method for 3D Mechanical Components Based on Object Projection, Pre-Trained Deep Learning Models and Autoencoder. Comput. Des. 154, 103417. https://doi.org/10.1016/j.cad.2022.103417

De Geyter, S., Vermandere, J., De Winter, H., Bassier, M., Vergauwen, M., 2022. Point Cloud Validation: On the Impact of Laser Scanning Technologies on the Semantic Segmentation for BIM Modeling and Evaluation. Remote Sens. 14. https://doi.org/10.3390/rs14030582

Frome, A., Huber, D., Kolluri, R., Büllow, T., Malik, J., 2004. Recognizing Objects in Range Data Using Regional Point Descriptors, in: Pajdla, T., Matas, J. (Eds.), Computer Vision - ECCV 2004. ECCV 2004. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 36–39. https://doi.org/10.1007/978-3-540-24672-5_18

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. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 11105–11114. https://doi.org/10.1109/CVPR42600.2020.01112

Iyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y., Ramani, K., 2005. Three-dimensional shape searching: State-of-the-art review and future trends. CAD Comput. Aided Des. 37, 509–530. https://doi.org/10.1016/j.cad.2004.07.002

Krishna, O., Irie, G., Wu, X., Kawanishi, T., Kashino, K., 2021. Adaptive Spotting: Deep Reinforcement Object Search in 3D Point Clouds. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 12624 LNCS, 257–272. https://doi.org/10.1007/978-3-030-69535-4_16

Li, B., Lu, Y., Li, C., Godil, A., Schreck, T., Aono, M., Burtscher, M., Chen, Q., Chowdhury, N.K., Fang, B., Fu, H., Furuya, T., Li, H., Liu, J., Johan, H., Kosaka, R., Koyanagi, H., Ohbuchi, R., Tatsuma, A., Wan, Y., Zhang, C., Zou, C., 2015. A comparison of 3D shape retrieval methods based on a large-scale benchmark supporting multimodal queries. Comput. Vis. Image Underst. 131, 1–27. https://doi.org/10.1016/j.cviu.2014.10.006

Li, J., Chen, B.M., Lee, G.H., 2018. SO-Net: Self-Organizing Network for Point Cloud Analysis. Cvpr2018 9397–9406.

Li, N., Li, Q., Liu, Y.S., Lu, W., Wang, W., 2020. BIMSeek++: Retrieving BIM components using similarity measurement of attributes. Comput. Ind. 116, 103186. https://doi.org/10.1016/j.compind.2020.103186

Liu, H., Xu, Y., Zhang, J., Zhu, J., Li, Y., Hoi, S.C.H., 2020. DeepFacade: A Deep Learning Approach to Facade Parsing with Symmetric Loss. IEEE Trans. Multimed. 22, 3153–3165. https://doi.org/10.1109/TMM.2020.2971431

Liu, Q., 2012. A Survey of Recent View-based 3D Model Retrieval Methods 1–15.

Long, J., Shelhamer, E., Darrell, T., 2017. Fully Convolutional Networks for Semantic Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 640–651. https://doi.org/10.1109/TPAMI.2016.2572683

Musialski, P., Wonka, P., Aliaga, D.G., Wimmer, M., Van Gool, L., Purgathofer, W., 2013. A survey of urban

reconstruction. Comput. Graph. Forum 32, 146–177. https://doi.org/10.1111/cgf.12077

Neuhausen, M., Koch, C., König, M., 2016. Image-based Window Detection - An Overview. 23rd Int. Work. Eur. Gr. Intell. Comput. Eng. EG-ICE 2016.

Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D., 2001. Matching 3D Models with Shape Distributions, in: Proceedings International Conference on Shape Modeling and Applications. pp. 154–166. https://doi.org/10.1109/SMA.2001.923386

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. PointNet: Deep learning on point sets for 3D classification and segmentation. Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 77–85. https://doi.org/10.1109/CVPR.2017.16

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. PointNet++: Deep hierarchical feature learning on point sets in a metric space. Adv. Neural Inf. Process. Syst. 2017-Decem, 5100–5109.

Quintana, B., Prieto, S.A., Adán, A., Bosché, F., 2018. Door detection in 3D coloured point clouds of indoor environments. Autom. Constr. 85, 146–166. https://doi.org/10.1016/j.autcon.2017.10.016

Robert, D., Vallet, B., Landrieu, L., 2022. Learning Multi-View Aggregation In the Wild for Large-Scale 3D Semantic Segmentation. IEEE/CVF Conference on Computer Vision and Pattern Recognition. https://doi.org/10.48550/ARXIV.2204.07548

Schmitz, M., Mayer, H., 2016. A convolutional network for semantic facade segmentation and interpretation. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 41, 709–715. https://doi.org/10.5194/isprsarchives-XLI-B3-709-2016

Sipiran, I., Bustos, B., Schreck, T., 2013. Data-aware 3D partitioning for generic shape retrieval. Comput. Graph. 37, 460–472. https://doi.org/10.1016/j.cag.2013.04.002

Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E., 2015. Multi-view convolutional neural networks for 3D shape recognition. Proc. IEEE Int. Conf. Comput. Vis. 2015 International Conference on Computer Vision, ICCV 2015, 945–953. https://doi.org/10.1109/ICCV.2015.114

Tan, M., Le, Q. V., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. 36th Int. Conf. Mach. Learn. ICML 2019 2019-June, 10691–10700.

Van Gool, L., Zeng, G., Van den Borre, F., Müller, P., 2007. TOWARDS MASS-PRODUCED BUILDING MODELS, in: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. Munich, Germany, pp. 209–220.

Xiao, Y., Lai, Y., Zhang, F., Li, C., Gao, L., 2020. A survey on deep geometry learning: From a representation perspective. Comput. Vis. Media 6, 113–133. https://doi.org/10.1007/s41095-020-0174-8

Zhou, Q.Y., Park, J., Koltun, V., 2016. Fast global registration. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9906 LNCS, 766–782. https://doi.org/10.1007/978-3-319-46475-6_47