SCAN-TO-BIM FOR SMALL BUILDING COMPONENTS

Blanca Quintana[1], Samuel A. Prieto[2], Antonio Adán[3] and Frédéric Bosché [4]

Abstract: Scan-to-BIM works have so far mainly focused on 'structural' components such as floors, ceiling, walls (with doors and windows). But, the control of new facilities and the production of their corresponding as-is BIM models requires the identification and inspection of numerous other building components and objects, e.g. MEP components such as plugs, switches, ducts, and signs. In this paper, we present a novel 6D-based (XYZ + RGB) approach that processes dense coloured 3D points provided by terrestrial laser scanners to recognize such smaller objects that are commonly located on walls. This paper focuses on the recognition of objects such as sockets, switches, signs, and extinguishers. After segmenting the point clouds corresponding to the walls of a building, a set of candidate objects are detected independently in the colour and geometric spaces, and a consensus procedure integrates both results to infer recognition. The method has been tested on real indoors yielding promising results.

Keywords: object recognition, scan-to-BIM, automatic BIM, 3D data processing.

1 BACKGROUND

Efforts in automatic Scan-to-BIM modelling from laser scanning or photogrammetric data have so far been focused on segmenting and recognizing large architectural 'structural' components like walls (with openings), floors, ceiling and columns. But, automatic building modelling from point cloud data needs to go beyond those large 'primary' components to other smaller 'secondary' ones. For example, walls do not contain just openings (doors and windows) but many other building service components such as sockets, safety lights or alarm devices, that should be included in as-is BIM models for effective O&M using BIM. In the remaining of this manuscript, we will refer to those objects as 'small components'.

The automatic detection and localization of small-objects in point cloud data is a challenging research line that has however received comparatively little research interest to date. Apart from the works that detect openings (door, windows) in walls using 2D and/or 3D information (Borgsen et al. 2014; Varadarajan & Vincze 2010; Yuan et al. 2015), only a few works deal with the recognition of other smaller, building service components like luminaries and sockets.

Focusing on lights and luminaries, Vilariño et al. (2015) distinguish fluorescent lighting tubes from circular low energy bulbs in 2D images of ceilings. In (Puente et al. 2014) another luminary detection approach is presented for application in tunnels.

[1] Research Assistant. Visual Computing and Robotic Lab. Castilla La Mancha University, Ciudad Real, Spain, blanca.quintana@uclm.es
[2] Research Assistant. Visual Computing and Robotic Lab. Castilla La Mancha University, Ciudad Real, Spain, samuel.prieto@uclm.es
[3] Associate Professor, Computer Science School, Leader of the Visual Computing and Robotic Lab. Castilla La Mancha University, Ciudad Real, Spain, antonio.adan@uclm.es
[4] Associate Professor, School of Energy, Geoscience, Infrastructure and Society, Leader of the CyberBuild Lab., Heriot-Watt University, Edinburgh, UK, f.n.bosche@hw.ac.uk

Regarding electrical components, in (Eruhimov & Meeussen 2011) electrical outlets are detected in coloured images acquired by a robot and classified as power holes, ground holes and background. Since the regions around the holes do not have any texture, they apply an intensity-based method for recognition. To calculate the 3D coordinates, a planar PnP solver is applied. In (Meeussen et al. 2010) a solution is proposed that recognizes doors, door handles, electrical plugs and sockets in an office environment using mobile robots. To detect door handles, digital images, range data and point clouds are processed separately. Outlet detection is carried out with disparity images, point clouds and colour images. In (Krispel et al. 2015), sockets and switches are recognized in orthographic 2D images. Object detection is performed with a simple sliding window and patch matching approach. Detection probability is measured using a feature descriptor pool (a kind of HoG that models the distribution of image gradients in different way) and a random forest classifier is used for classifying detections as power socket, light switch or background. Kang et al. (2010) detect elevator call buttons using range data coming from a stereo camera. First, an adaptive thresholding generates a binary image of the elevator. Then, some buttons and the floor number candidates are found in this image. Ambiguous candidates are rejected using an artificial neural network, and finally a matching method is applied to recognize, not only the call buttons, but more properties like destination floor buttons, moving direction and current location of the elevator. Czerniawski et al. (2016) recognize pipe spools in cluttered point cloud scans. The method is based on local data level curvature estimation, clustering and bag-of-features matching.

An interesting work is that of Hamledari et al. (2017). The proposed algorithm process 2D images and detects insulation, studs, electrical outlets and different states for drywall sheets. This information could provide valuable information for progress tracking systems. Focusing on different objects, Bonanni et al. (2013) detect: fire extinguishers, hydrant boxes and printers. In that work, the system requires human intervention. After the robot has acquired an 6D (XYZ-RGB) image of the scene, object segmentation requires the human to point out, with a laser pointer, the object to be segmented

2   PROPOSED APPROACH

We propose a new system for the detection of 'small components' in coloured point clouds acquired by a 3D laser scanner calibrated with a digital camera, altogether mounted on board a mobile robot. The robot goes through the rooms of a building and collects various scans that are automatically co-registered in the same coordinate system. The subsequent object recognition stages are described in the rest of this section.

2.1   Prior Data Processing/Preparation: Detection of Wall Area and Openings

The primary 'structural' components of the interior environment, i.e. walls, ceiling, floor and columns are first detected in the data and modelled – see (Quintana et al. 2016) for details. The input of our object recognition system is any coloured point cloud associated to a detected and modelled wall. More precisely, we generate a 4D orthoimage of the wall, J_{CD}, where each pixel has colour (RGB) and depth (i.e. orthonormal distance of the 3D points to the wall plane). The resolution of the 4D RGB-D orthoimage is 5mm × 5mm per pixel.

Since different light sources may impact a wall from different points of the room, slight colour variations (with boundaries) usually appear in the visible wall areas. Consequently, simple colour thresholding algorithms over our orthoimage would likely not be effective

enough for colour-based object detection. We therefore pre-process the 4D orthoimage to reduce such colour variations. Additional, specular highlights resulting from the system's camera flash are automatically detected and corrected. We refer the reader to (Quintana et al. 2016b) for details.

The wall 4D orthoimage is then processed in two stages. First, we detect the visible areas of the walls, as well as the openings (windows and doors), as fully explained in (Quintana et al. 2016b). The second stage consists in recognising and modelling small-objects such as switches, plugs and signs. This paper focuses on this latter stage.

Our approach assumes the existence of a database of objects that may be present in the data. This database includes, for each object, colour and the depth image models. Note that, in the case of a Scan-vs-BIM scenario, this database could be automatically generated using the as-design BIM model of the facility. The as-design BIM model would also provide additional valuable information, such as which objects should be present in that wall. In this paper, we very much put ourselves in such a context.

Our small component recognition approach then processes the geometric (i.e. XYZ coordinates) and colour (i.e. RGB) information separately (Sections 2.2 and 2.3) and combines their results in a consensus decision making approach (Section 2.4).

2.2 Small Component Recognition with geometry

After detecting and positioning the doors (or windows) in the orthoimage J_{CD} , we generate a new 4D orthoimage $\hat{J}_{CD}$ that exclusively contains the information contained in the wall area, and $\hat{J}_{CD}$ is again decomposed into a depth image $\hat{J}_D$ and a colour image $\hat{J}_C$. $\hat{J}_D$ is used to detect objects with geometric discontinuities with respect to the wall plane, which is valuable for the detection of protruding objects; $\hat{J}_C$ is used to detect objects with colour discontinuities with respect to the wall area.

A decreasing Canny filter is iteratively applied to $\hat{J}_D$, obtaining, for the kth iteration, the image $\hat{J}^k_{D,Canny}$. The Canny's threshold range, which goes from 1 to 0, is decreased in order to find salient regions in the wall. For the kth iteration, the regions corresponding to the bounding boxes of the candidate salient objects in earlier iterations $\{k-1, k-2, ...,1\}$ are removed from the current $\hat{J}^k_{D,Canny}$. The resulting image, $\widehat{J}'^k_{D,Canny}$, is then processed to detect closed boundary regions, each representing a candidate to be further investigated. Obviously, as the Canny threshold decreases, the image $\widehat{J}'^k_{D,Canny}$ has got fewer salient objects. When the Canny threshold is below 0.05, the iterative process is stopped and the candidate database is analysed.

The list of candidate regions is first reduced by clustering fully overlapping bounding boxes, with the enclosing bounding box selected as the cluster's prototype candidate. Then, a simple image cross-correlation algorithm is applied that assesses the correlation between the candidate regions with depth image models contained in the given database of objects to be recognised. Cross-correlation is applied, not on the depth images directly, but on the corresponding gradient images. Note that both the input 4D orthoimage, and therefore $\hat{J}_D$, and the database depth image models have known scales; therefore, it is not necessary to conduct the cross-correlation operations at multiple scales. The cross-correlation function also yields the position of the recognized object in $\hat{J}_D$.
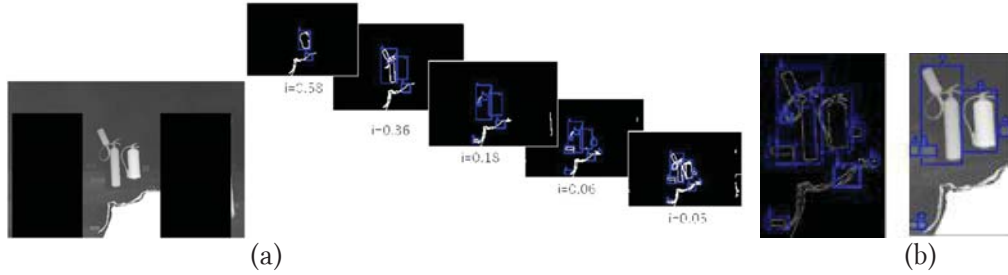
Figure 1: (a) Images $\hat{J}_D$ and $\widehat{J'}^{ik}_{D,Canny}$ after several iterations (i is the Canny threshold); (b) Detail of the candidate regions and recognized objects in $\hat{J}_D$

2.3   Small Object Recognition with colour

Objects that are salient in the colour domain instead of the geometric domain are detected by looking for colour discontinuities in $\hat{J}_C$. Firstly, we discard the visible wall area from $\hat{J}_C$, so that it contains only parts of the wall that are neither the visible wall area nor openings. Then, the resulting image, $\widehat{J'}_C$, is transformed into a binary image $\hat{J}_{BW}$ (as illustrated in Figure 2), that is processed with the objective of finding compact sets of pixels, each theoretically representing a candidate region for object detection. Similarly to the recognition approach applied to the depth image, the bounding box is calculated for each candidate region and the corresponding region in $\widehat{J'}_C$ is extracted and stored.

The matching against objects in the database established a priori is conducted by matching SURF features (Bay et al. 2006) extracted from the colour image models in the database with those extracted from the candidate region. Figure 2c shows image $\hat{J}_{BW}$ and the initial set of bounding boxes calculated in the segmentation process.
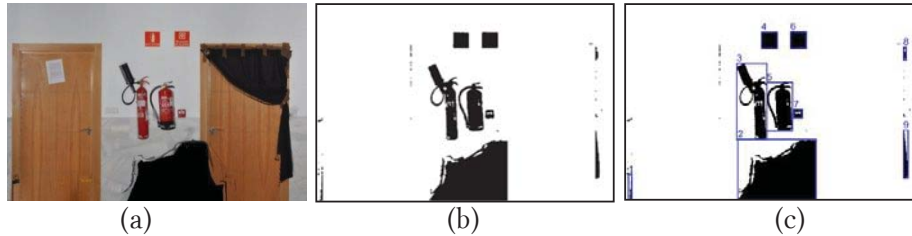

Figure 2: (a) Image J_C, (b) image $\hat{J}_{BW}$ and (c) the initial set of bounding boxes calculated in the segmentation process.

2.4   Consensus Strategy for Recognition and Positioning

Although some objects might be detected through both geometry and colour, there are some others that may be recognized in only one of the datasets. Furthermore, several instances of a database object may be present in the wall. To find a consensus between depth-based and colour-detections and address the potential presence of multiple instances, the following method is proposed.

For each object in the database, we collect all its detection candidates from $\hat{J}_D$ and $\hat{J}_C$, and construct a *Recognition Coherence Matrix*, R, as illustrated in Figure 3. Each entry in R measures the coherence between a detection region in $\hat{J}_D$ (or none) and another in $\hat{J}_C$ (or none). The *Recognition Coherence Level* $\alpha \in [0,1]$ between two regions is calculated by assessing the overlap between the two region bounding boxes B_C^i and B_D^j (with centroid coordinates c_C^i and c_D^j respectively), as:

$$
\alpha = \begin{cases}
(I) & \dfrac{\langle B_C^i \rangle \cup \langle B_D^j \rangle}{\langle B_{CD}^{ij} \rangle} & if \ B_C^i \cap B_D^j \neq \emptyset \\[2ex]
(II) & 0 & if \ B_C^i \cap B_D^j = \emptyset \\[2ex]
(III) & 0.5 & if \ \nexists B_C^i or \nexists B_D^j
\end{cases} \tag{1}
$$

where B_{CD}^{ij} is the bounding box that encloses B_C^i and B_D^j, and $\langle B \rangle$ is the number of pixels covered by the bounding box B.

If the bounding boxes are overlapping, α takes high values (next to 1), otherwise α progressively decreases as the distance grows (case I). If the bounding boxes are not overlapping $\alpha = 0$ (case II). For cases where the database object is detected in only one of the two images or it does not match favourably with any detection in the second image, we set α as 0.5 (case (III).

Once R is filled, the recognition's consensus decision is solved iteratively for each entry detection candidate, as follows (see Figure 3b): (1) the highest value of R is selected and considered to be a recognised instance of the database object; (2) the corresponding row and column of R are eliminated and the process iterated until R is null.

The location (centre) of each recognised object instance, c, is then calculated using equation (2):

$$
c = \frac{\rho_D c_D + \rho_c c_C}{\rho_D + \rho_C} \tag{2}
$$

where the two weights ρ_D and ρ_C are the cross correlation coefficient (which evaluates the goodness of the recognition in the depth image) and the SURF SSD mean (which evaluates the goodness of the recognition in the colour image).

(a)   (b)

Figure 3: a) Example of *Recognition Coherence Matrix* and b) Stages of the recognition's consensus decision process.

3   EXPERIMENTAL TEST

Our small building component recognition approach has been tested in real coloured point clouds collected from a Riegl VZ400 laser scanner associated to a DLSR camera within a robotic platform. In this section, we analyse the results obtained in a representative case of study. The orthoimage J_{CD} of Figure 4 corresponds to one of the walls of the Computer Science School of Castilla La Mancha University. J_{CD} was obtained after scanning the corridor from three different locations. This is a good and complex case study for several reasons. First, the wall contains several objects which can only be detected by geometry

or colour. Second, a part of the wall has been tiled with a kind of reflecting tiles, which introduces noise in the depth data and makes the processing more difficult.

In this test, we assume that the objects to be recognised have been obtained from the as-designed BIM model of the building, and we are looking to create an as-is model of the facility from the scanned data. The database is elaborated with the most representative small components lying on the walls of the building's as-designed BIM model. The geometric models (depth images) are built in our lab by using a Minolta 910 laser scanner and a turntable. The colour models are extracted from public websites. As can be seen in Figure 4, the wall contains the following objects: two socket blocks (one with one socket and one with three), two switch blocks (three with one switch and one with three), two extinguishers, one fire-alarm box and two alarm signs.

Table 1 presents the recognition and positioning results of the test. The best values of parameters ρ_D and ρ_C are in the third and fourth columns. The values of ρ_D and ρ_C must exceed the threshold 0.55 to consider the detection as a potential candidate recognition (i.e. to be added to R). The highest values of α are achieved in two cases (Extinguisher 1 and Extinguisher 2) where the object is recognized in both $\hat{J}_D$ and $\hat{J}_C$. Five objects were exclusively recognized using geometry, and two were recognised exclusively using colour. The objects #2 and #4 are not recognized. As a result, the recall rate is 81,8%.

Regarding the localisation accuracy, columns entitled Δ_1 and Δ_2 provide the horizontal and vertical positioning error (in millimetres) of the recognized object's centre within the wall plane, for which the ground truth was manually measured. In general, errors are below 10 mm, which can be considered an accurate positioning result. Figure 4 illustrates some of the results and Figure 5 shows the integration of both approaches.

Table 1: Results of the experimental test

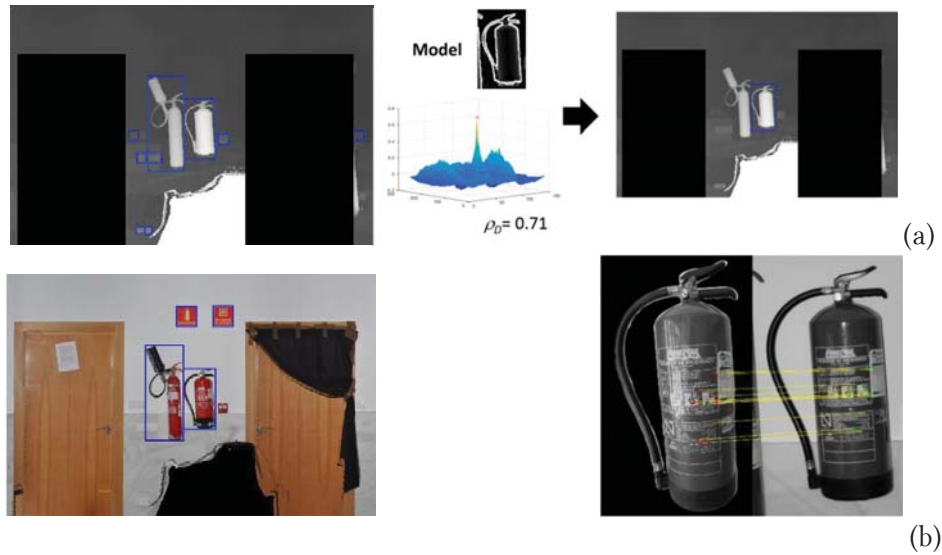| Item #	Object	ρ_D	ρ_C	α	Δ_1	Δ_2	Recognized
1	Socket (3)	0,64	-	0.5	5	5	Yes
2	Socket	0,66	-	0	-	-	No
3	Switch (3)	0,58	-	0.5	5	5	Yes
4	Switch	0,57	-	0	-	-	No
5	Switch	0,77	-	0.5	0	5	Yes
6	Switch	0,75	-	0.5	0	5	Yes
7	Extinguisher 1	0,65	0.9510	0.99	0	10	Yes
8	Extinguisher 2	0,71	0.9583	0.99	2,5	2,5	Yes
9	Fire-alarm	0,72	-	0.5	5	0	Yes
10	Extinguisher-sign	-	0.9721	0.5	2,5	0	Yes
11	Alarm-sign	-	0.9664	0.5	2,5	2,5	Yes

(a)

(b)

Figure 4: a) Recognition results in $\hat{J}_D$. Rectangles in blue are obtained after finishing the iterative Canny's process. On the right, a detail of the cross correlation map of extinguisher 2 is illustrated. b) Recognition results in $\hat{J}_C$. On the right, a detail of the SURF algorithm over one of the extinguishers.

(a)   (b)

Figure 5: Integration of recognition results. a) Recognized objects by geometry (in cyan) and by colour (in magenta). b) True positives (in green), false positives (in red) and false negatives (in blue).

4   CONCLUSIONS

The fusion of imaged and geometric algorithms is a new strategy in the automatic creation of BIM models. This paper presents a novel 6D-based (3D coordinates + RGB) approach that processes dense coloured 3D points with the aim of recognizing small components in buildings. This approach is more effective and robust than others because some small components, that are not detectable in colour images (for example white sockets inserted in white walls), can be recognized in the geometric space and vice versa. Thus, our algorithm is able to recognize objects like sockets, switches, signs, alarm devices or extinguishers in walls. Future work is addressed towards recognition and positioning with wider object databases and in more complex scenes.

5   ACKNOWLEDGMENTS

6   REFERENCES

Borgsen, S.M.Z. et al., (2014). Automated door detection with a 3D-sensor. *Proceedings - Conference on Computer and Robot Vision, CRV*, pp.276–282

Varadarajan, K.M. & Vincze, M., (2010). 3D room modeling and doorway detection from indoor stereo imagery using feature guided piecewise depth diffusion. *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pp.2758–2765.

Yuan, T.H. et al., (2015). An Automated 3D Scanning Algorithm using Depth Cameras for Door Detection. *International Electronics Symposium (IES)*, pp.58–61

Díaz-Vilariño, L. et al., (2015). Automatic LiDAR-based lighting inventory in buildings. *Measurement: Journal of the International Measurement Confederation*, 73, pp.544–550

Puente, I. et al., (2014). Automatic detection of road tunnel luminaires using a mobile LiDAR system. *Measurement: Journal of the International Measurement Confederation*, 47(1), pp.569–575.

Eruhimov, V. & Meeussen, W., (2011). Outlet detection and pose estimation for robot continuous operation. *IEEE International Conference on Intelligent Robots and Systems*, pp.2941–2946

Meeussen, W. et al., (2010). Autonomous door opening and plugging in with a personal robot. *IEEE International Conference on Robotics and Automation.*

Krispel, U. et al., (2015). Automatic texture and orthophoto generation from registered panoramic views. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(5W4), pp.131–137.

Kang, J.G. et al., (2010). Recognition and path planning strategy for autonomous navigation in the elevator environment. *International Journal of Control, Automation and Systems,* 8(4), pp.808–821.

Czerniawski T, Nahangi M., Haas C., Walbridge S. Pipe, (2016). Spool recognition in cluttered point clouds using a curvature-based shape descriptor. *Automation in Construction.* Vol 71, Part 2, Pages 346–358.

Hamledari H, McCabe B., Davari S (2017). Automated computer vision-based detection of components of under-construction indoor partitions. *Automation in Construction*, Vol 74, February 2017, Pages 78–94.

Bonanni, T.M. et al., (2013). Human-Robot Collaboration for Semantic Labeling of the Environment., *Cognitive Cooperating Robots,* pp.1–25.

Quintana, B., Prieto, S.A., Adán, A. & Vázquez, A.S., (2016). Semantic Scan Planning for Indoor Structural Elements of Buildings. *Advanced Engineering Informatics.* Vol 30, Issue 4, October 2016, Pages 643–659.

Quintana, B., Prieto, S.A., Adán, A. & Bosché, F., (2016). Door Detection in 3D Colored Laser Scans for Autonomous Indoor Navigation. *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, (October).

Bay, H., Tuytelaars, T. & Van Gool, L., (2006). SURF: Speeded up robust features. *Lecture Notes in Computer Science*, 3951 LNCS, pp.404–417.