Automated Topometric Graph Generation from Floor Plan Analysis

Obadiah Lam, Feras Dayoub, Ruth Schulz, Peter Corke[1]
ARC Centre of Excellence for Robotic Vision, Queensland University of Technology (QUT) * †

Abstract

The world is rich with information such as signage and maps to assist humans to navigate. We present a method to extract topological spatial information from a generic bitmap floor plan and build a topometric graph that can be used by a mobile robot for tasks such as path planning and guided exploration. The algorithm first detects and extracts text in an image of the floor plan. Using the locations of the extracted text, flood fill is used to find the rooms and hallways. Doors are found by matching SURF features and these form the connections between rooms, which are the edges of the topological graph. Our system is able to automatically detect doors and differentiate between hallways and rooms, which is important for effective navigation. We show that our method can extract a topometric graph from a floor plan and is robust against ambiguous cases most commonly seen in floor plans including elevators and stairwells.

1 Introduction

Our world is rich with information such as signage and maps that are explicitly created to assist humans in navigation, but much of this is currently inaccessible to robots. This has led to a significant body of research in robotic mapping and exploration, yet in many cases the information obtained already existed, just in a form that robots could not easily access. For example, modern buildings are built to an architectural floor plan, typically using a CAD system. The original CAD files contain a wealth of spatial information but are not generally available, however PDF versions of floor plans for many buildings are available online and evacuation plans are available on the walls of many buildings. This paper presents an automated technique to extract room labels and topometric structure from such bitmap image floor plans.

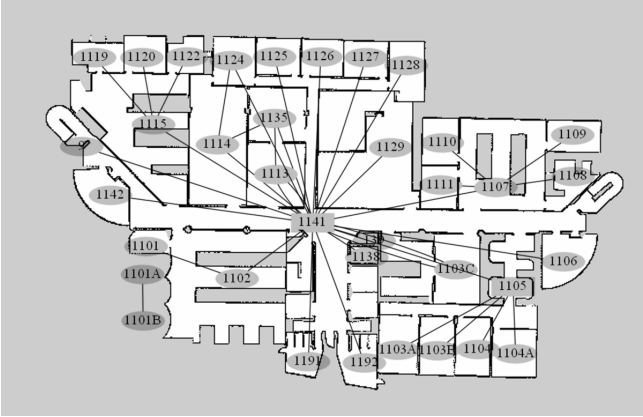


Figure 1: The proposed system takes as an input a bitmap image of a floor plan and returns a topometric map with the rooms as nodes and the doors as edges, shown here overlaid on an occupancy grid map.

In a typical floor plan, spaces (such as rooms and hallways) are assigned semantic labels such as room numbers or names, which are included on the floor plan as text within the space they refer to. These semantic labels are generally reflected in the physical world by door labels and signs. Using a text recognition system, a robot can localise itself in the map by reading such labels and signs. Room label and location information from a floor plan was also used in [Schulz *et al.*, 2015] to perform goal-directed exploration in unfamiliar environments. While that information was extracted manually from the floor plan, the methods in this paper will automate that process.

*This research was supported under Australian Researchz Council's Discovery Projects funding scheme (project number DP140103216)

†[1]The authors are with the ARC Centre of Excellence for Robotic Vision, Queensland University of Technology (QUT), Brisbane, Australia. http://www.roboticvision.org/ email: {o1.lam, feras.dayoub, ruth.schulz, peter.corke}@qut.edu.au

The contributions of this paper include the first application of bitmap floor plan analysis to robotics and the automatic generation of a topological graph. The nodes in this graph are labelled with human-interpretable semantic labels such as room labels, as shown in Figure 1. These could be used to match with the results of an internet search, for example. We also combine both coarse topological information and metric information from the floor plan to generate a topometric graph. This graph is embedded in robot-friendly metric coordinates, not to scale but in [Schulz *et al.*, 2015] we have shown that scale can be estimated by observing landmarks.

This paper extends the text spotting work in [Posner *et al.*, 2010] and [Lam *et al.*, 2014] by adding an additional source of prior information that can be used for robot navigation, and continues to investigate the assignment of meaningful human semantics to space. The methods and results from this work have been used in goal-directed robotic navigation, but can also be used for applications such as path planning for walking directions [Whiting *et al.*, 2007].

The rest of the paper is organised as follows. In Section 2, we discuss related work in the field. Section 3 gives an overview of our algorithm. We describe an experiment to test the performance of this system in Section 4. Finally Section 5 presents the results of applying our method to floor plans from varying sources and we draw conclusions and discuss future work in Section 6.

2 Related Work

Within the robotics community, research has already been undertaken into robot navigation using abstract sources of spatial relationships. Luo *et al.* [Luo *et al.*, 2010] developed a method of robot navigation and localisation using floor plan information by combining Support Vector Machines to read text labels and detecting passage corner landmarks using ultrasonic sensors. Floor plans have also been used to build occupancy grid maps, where localisation is then performed with a depth sensor and WiFi [Ito *et al.*, 2014], but this research focuses on the localisation problem without extracting any useful semantic information from the floor plan. The inverse process has also been investigated, where an occupancy grid map is used to extract spatial information and build an abstracted floor plan [Liu *et al.*, 2012].

Topological and hierarchical maps built from metric robot maps can also be combined with semantic information about the spatial layout as a robot moves through the map. This may involve vision approaches [Galindo *et al.*, 2005] or feature boosting for a classifier of both vision and laser scan data [Mozos *et al.*, 2007]. These approaches rely on the existence of a metric occupancy grid map.

Notably, there has also been work in understanding trends in spatial layouts from analysing a large corpus of floor plans. Aydemir's work in [Aydemir *et al.*, 2012] analysed indoor environments using large floor plan datasets and found that local structure of an indoor environment was independent of the global structure and was therefore predictable. These analyses and results were then used to predict the topological structure of yet unexplored areas in the environment to augment a robot's understanding of its surroundings. However, the floor plans are represented in an XML format where room centroid coordinates, labels and spatial relationships including doorway locations are already explicitly represented. This is unlike our method, which extracts this information directly from the graphical representation (i.e. bitmap) of the floor plan.

Within the document analysis community there has been some related work in floor plan bitmap image analysis. This includes [Macé *et al.*, 2010] where text labels are assumed to be absent and the Hough Transform is used, coupled with image vectorisation. Doors are detected with arcs and walls are detected from straight lines. Ahmed *et al.* have a similar system for automated floor plan analysis [Ahmed *et al.*, 2011] by performing wall detection using thick/thin line separation, relying on the assumption that the rooms and spaces are rectangular in shape. Our method does not rely on this assumption and will work for rooms and spaces of any shape.

The spatial information from floor plans has been used in other applications, such as planning walking directions between two points on a university campus [Whiting *et al.*, 2007]. Results show that this spatial information improves route planning by employing nearest building entrances rather than street-facing entries and could lead to further work in understanding human behaviour and navigation in urban environments. Similar techniques have successfully been applied to navigational aids for visually impaired persons in [Joseph *et al.*, 2013a], where an augmented reality system directs the user with a haptic belt and voice guidance [Joseph *et al.*, 2013b].

Our work proposes a method of extracting this valuable spatial information from the bitmap image of the floor plan itself. Both humans and robots can use the topometric graph built from this spatial information to understand an indoor space. It differs from existing document analysis research as we do not rely on wall detection with line extraction. Our method is thus more robust against the varying architectural standards and visual representations (where a wall may be represented by a single thick line, multiple thin lines, etc.)

3 Approach

In this section the method for generating the topological and topometric graphs is outlined. First, the input

file (PDF or image) is converted to a binary bitmap image. Next, the text and room labels are detected and extracted from the resultant image, as briefly described in Sec. 3.1.

The topological graph is formed with the rooms and hallways as nodes and the doorways connecting these spaces as the edges. The door detection process is outlined in Sec. 3.2 and the formation of the spatial regions that are the room hypotheses is described in Sec. 3.3.

3.1 Text Extraction

A complete solution to the text detection and extraction problem is not the focus of this paper. Existing methods of text detection are implemented. Some of the 'characterness' cues outlined in [Li *et al.*, 2014] are used to detect and subsequently extract the text from the floor plan image. Specifically, Minimally Stable Extremal Regions (MSER) [Matas *et al.*, 2004] is used as a region detector followed by the Stroke Width Transform [Epshtein *et al.*, 2010]. We skeletonise the regions to reduce computation time, using a distance transform to find the stroke width at those pixels. A region is considered a text character if the stroke width is consistent across the whole skeleton. Weak geometric constraints such as aspect ratio are applied to filter out noisy regions, such as long, thin lines. The pixels that are detected as text are then removed from the image.

Each MSER text region is dilated using a horizontal bar structuring element to collect characters into words. Using the bounding box of the word, the text region is cropped from the original image. Tesseract [Smith, 2007] is then used to perform text recognition on these cropped regions.

3.2 Door Detection

In our topological graph the doors form the edges between nodes. There are common symbols and abbreviations found on floor plans [Koel, 1999]. We use the standard symbol for interior doors as a template, seen in Figure 2. Local feature descriptors are used as an arc detector in [Ahmed *et al.*, 2011]. We chose Speeded Up Robust Features (SURF) [Bay *et al.*, 2008] for robustness and invariance against translation, rotation and scale changes. Door symbols in the map at any orientation can be detected with a single template without inverting or rotating the template, including double doors. False positives are also rejected, including curved outer walls and stairwells.

First, the key points are extracted from both the template door image and the floor plan image. The SURF descriptor is then computed for these key points. All detected SURF features from the floor plan image are compared against the strongest features in the template. A detected SURF feature is matched if the match distance

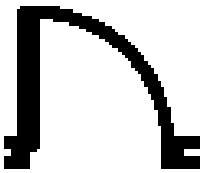


Figure 2: Single Door Template

to the closest template feature is less than a threshold λ. Figure 3a shows all the matching keypoints highlighted in a small area of a floor plan.

Matching SURF features are then collected using distance-based geometric clustering with a distance threshold d. This clearly depends on scale which we can assume is known a priori or from the scale of the SURF keypoint descriptors. Each cluster of keypoints is labelled as an individual door and the midpoint of all the keypoints within the cluster is taken as the location of the door. The detected doors in the same small area of the floor plan are shown in Figure 3b.

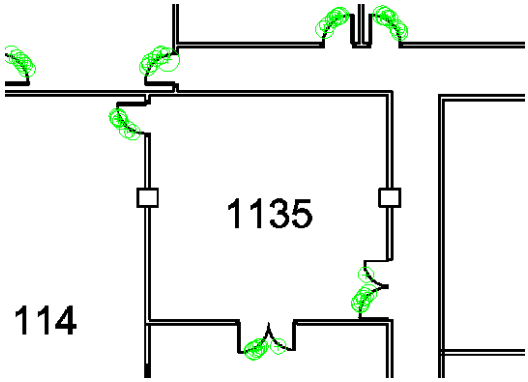
3.3 Room Detection

To divide the floor plan into its constituent rooms, a flood fill algorithm is used. First, the detected text is removed from the image. Pixels on the detected text are used as the seed points of the flood fill algorithm. In this regard the room detection depends on the text extraction stage and the assumption that the text label lies within the room on the floor plan. An assumption is also made that each room space is fully enclosed by black pixels, which tends to hold except in cases of sliding doors. For visualisation purposes, the outputs of the flood fill operation for room detection are colourmapped randomly. Figure 3c shows detected and flood filled rooms from the same small area of the floor plan.

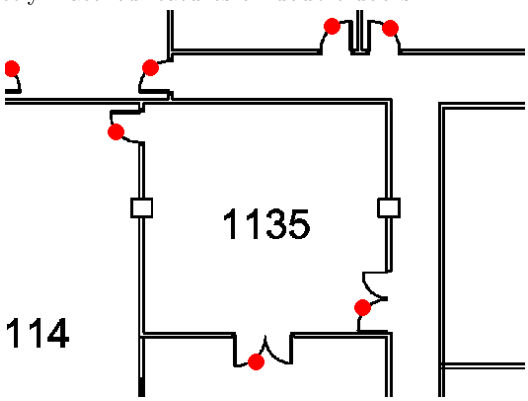
3.4 Topological/Topometric Graph

A door is considered to be the connection between two rooms, with spaces connected by sliding partitions and open plan areas considered to be one room. These standard doors form the edges between the nodes on our topological graph. The rooms that are connected by each door are determined by examining the output from the flood fill operation in 3.3.

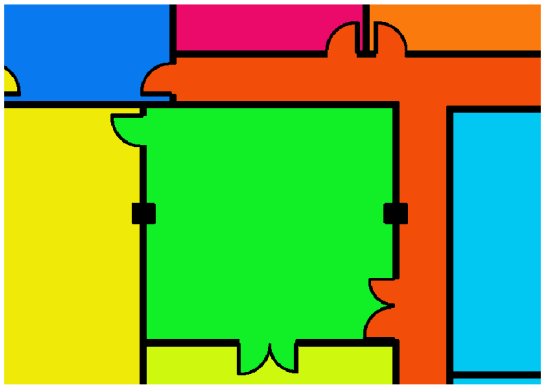
First, from each door location we search for the closest black pixel. This is taken to be a door pixel. From this door pixel a search for the nearest white room pixel is performed. This is taken to be the first room. That room is then temporarily flood filled to black using the white pixel that was found as the seed point. Another search is performed and the new nearest white room pixel is the connecting room.



(a) Matched SURF features to door template. Note correctly matched features on double doors.



(b) Detected Doors



(c) Detected Rooms, flood filled and colour mapped for visualisation

Figure 3: Door and Room Detection

It is semantically useful to distinguish between rooms and hallways. For planning purposes a robot might need to use a hallway to move from a room to another. This also provides context for vision, such as scene classification or text recognition; for example, hallway doors are often labelled. Some basic geometric constraints are used to distinguish a discrete space as either a room or a hallway. Hallways tend to be elongated, with a large perimeter compared to their areas. This is a more effective cue than merely counting the doors leading into or out of a space. To measure the perimeter to area ratio, we use:

$$Q = \frac{4\pi A}{L^2} \tag{1}$$

where Q is the isoperimetric quotient, A is the area of the space and L is the perimeter. The aspect ratio is measured by calculating the fill factor of the space with respect to the squared bounding box:

$$F = \frac{A}{(\max(w,h))^2} \tag{2}$$

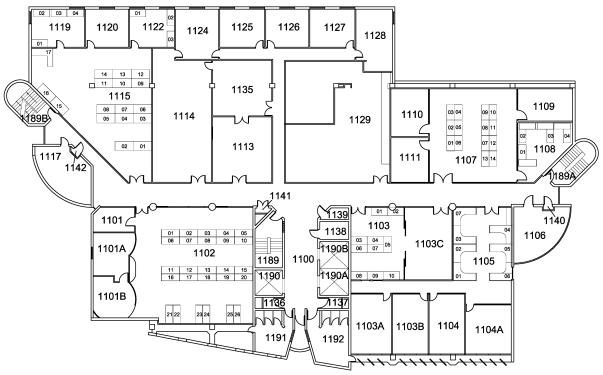
where F is the fill factor, A is the area of the space and w and h are the width and height of the space respectively.

A space is considered a hallway if the isoperimetric quotient Q is less than some threshold γ and the fill factor F is less than some other threshold δ. The fill factor F is used as an additional measure to further separate hallways from spaces that have a larger perimeter from complicated, non-smooth architectural structures such as windows. Note that merely counting edges into a node is insufficient to distinguish rooms from hallways as seen Figure 6c. Room 1115 has four connecting edges and hallway 1105 has five.

This information is sufficient to create a topological graph which the robot can navigate over using an algorithm such as A* [Hart *et al.*, 1968]. We visualise the graph using Graphviz [Gansner and North, 2000], as can be seen in Figure 1. The location of extracted text is used to assign labels to each room or hallway. The centroids of this text are used as metric information to position the nodes in the topometric graph. Different node symbols are used to differentiate hallways from rooms. Here, the rectangular nodes are hallways where the elliptical nodes are rooms. In some floor plans there are multiple lines of text within a room, either by design or when a room label runs outside of its room on the floor plan. We account for this by selecting the line of text closest to the centroid of the detected room.

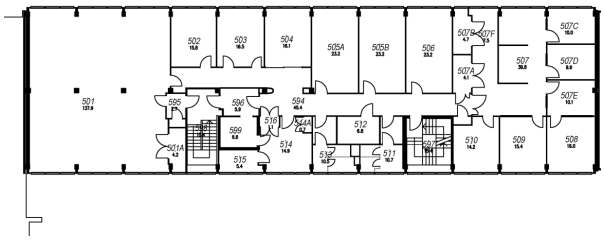
4 Experiment

We first investigate the performance of our algorithm against the quality of the bitmap floor plan, represented by the resolution of the image in pixels per inch (ppi).



(a) QUT S Block level 11

(b) UQ GP South level 7



(c) UQ Axon Building level 5

Figure 4: Input Floor Plans for Testing

The PDF of the floor plan was converted into bitmaps at five different resolutions: 72ppi, 96ppi, 150ppi, 300ppi and 600ppi. These were then resized to the same image dimensions, which in this case were 4963x3509 pixels. We use the QUT S Block architectural floor plan (shown in Figure 4a). Our floor plan analysis algorithm is then applied to these bitmap images.

The performance measure used is the number of doors correctly detected out of the ground truth doors, as well as the number of rooms and hallways correctly detected and filled out of the ground truth. Detected and recognised room labels are compared to the ground truth, though these only affect the labelling of the nodes in the constructed topological graph.

We also compare the robustness of our algorithm across floor plans in different styles from different architects and locations. Our algorithm is tested on the following three floor plans at 300dpi:

- Queensland University of Technology Gardens Point S Block Level 11 (Figure 4a)
- University of Queensland General Purpose South Level 7 (Figure 4b)
- University of Queensland Axon Building Level 5 (Figure 4c)

We use the same performance measures as above.

For both tests above we use the following parameters outlined in Table 1. We use 25% similarity for both door detection and room/hallway segmentation.

5 Results and Discussion

The results from the first test is shown in Figure 5. The reduced information in the lower resolution images affects the door detection accuracy more than the room detection accuracy. Both the door and the room detection accuracy stop rising around 300ppi resolution, with no improvement in performance above this resolution. At 600ppi, the size of the input image increases to 69.7 megapixels. Computational power is a limiting factor at such high resolutions. This gives an indication of the quality of image that will be required if images of campus maps captured by a robot's camera are to be processed. Improvements to the text extraction and door recognition stages may reduce the required image quality.

The results from the three different floor plans are shown in Table 2. Figure 6 shows each step of our approach, applied to the QUT floor plan, and Figure 7

Table 1: Parameters

Parameter	Value
λ	0.25
γ, δ	0.25
d	35 pixels (half the width of a door)

Table 2: Performance across three different floor plans at 300dpi [Correctly Detected and Recognised/Total (Accuracy)]

Floor Plan	Doors	Rooms	Room Labels
QUT S L11	49/54 (0.91)	38/47 (0.81)	32/36 (0.89)
UQ Axon L5	33/41 (0.80)	27/30 (0.90)	13/22 (0.59)
UQ GP S L7	33/55 (0.60)	48/59 (0.81)	19/29 (0.66)

Figure 5: Plot of room and door detection accuracy against bitmap resolution

shows the final result on the other two floor plans. Across all three floor plans, the room detection rate is at least 80%. However, the door detection accuracy is low for the UQ GP South floor plan at only 60%. This is due to a number of reasons and this is discussed further.

The text labels were extracted from the images with Tesseract, which did not correctly read 100% of the text (refer to Table 2). This is often due to occlusions of the text in the floor plan, or when two room labels actually overlap each other, as seen in Figure 8a. The quality and font of the text also affects Tesseract's ability to correctly read the text labels. Figure 8b shows an example where the room was not correctly labelled because the room label was not physically within the room. Conversely, there are multiple pieces of text in most of the rooms in the UQ floor plans.

Due to these imperfections in the text extraction, there are spaces that are not correctly flood filled. This includes elevators and stairwells, as well as some smaller rooms such as service closets. This could be overcome by improvements in the text detection and extraction, as well as the implementation of thin line removal as in [Dosch *et al.*, 2000], and will be the focus of future work. However, those failures do not significantly affect the overall topological graph nor its utility for visiting non-utility rooms. Only the spaces that are not correctly labelled are affected. These are simply not included in the graph. A robot using this graph would require additional information if the goal node was not included in the graph, or if it needed to plan a path through an incorrectly labelled node.

Enclosed spaces with no doors are not included in the graph. This removes structures such as numbered desks. Our door detection fails when the door in the image looks markedly different from the template. This is seen in Figure 8c, which shows a sliding door. The door detection rate is also lower in the GP South floor plan due to the curvature of the walls, which skews the door symbols. Figure 8d is an example of a tight physical space where the door arcs overlap. The difficulty of this area is compounded by the room labels, which overlap both the doors themselves as well as the walls. Without proper context, it is difficult even for humans to correctly read and identify these room labels.

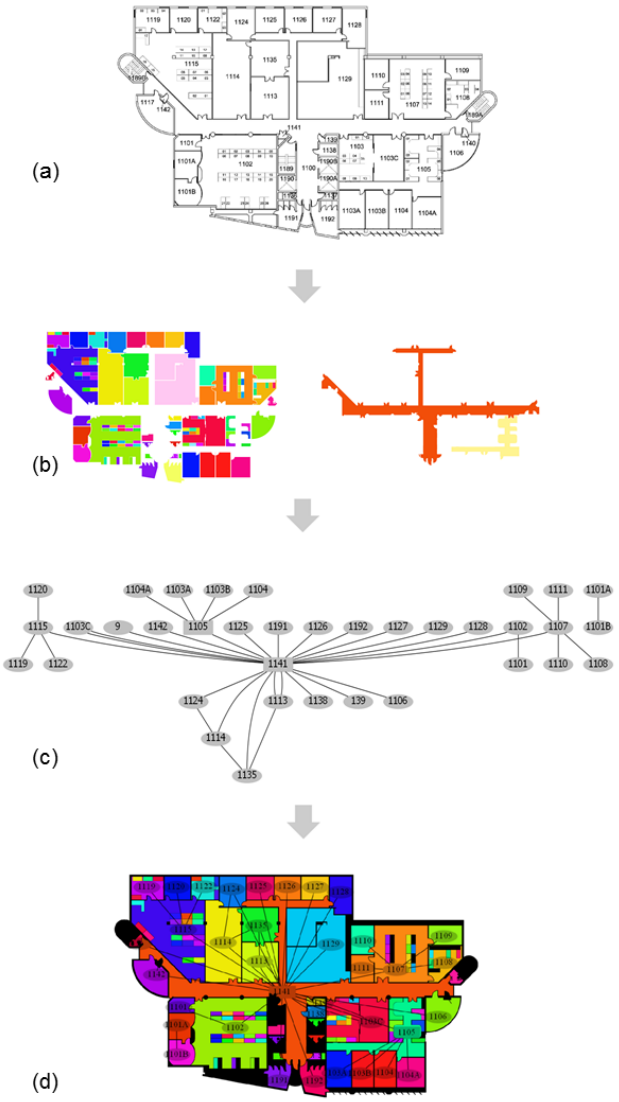
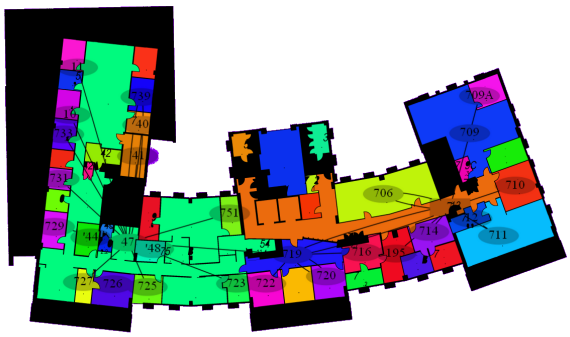


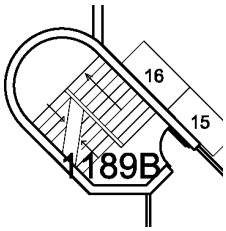
Figure 6: Results of our algorithm, applied to QUT S Block Level 11 floor plan (a). First, rooms are detected in (b), which form the nodes in the topological graph in (c). We embed this in metric space using the locations of detected room label text and overlay it on the colourmapped rooms in (d).

(a) UQ Axon Building Level 5

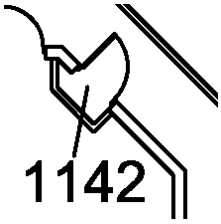


(b) UQ GP South Level 7

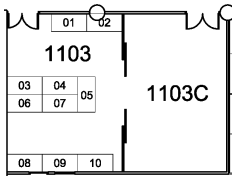
Figure 7: Topometric Map overlaid onto colourmapped floor plans



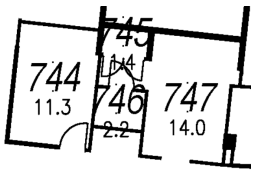
(a) Occlusions of Room Label



(b) Room Label not physically within room



(c) Sliding Doors



(d) Overlap of Door Arcs

Figure 8: Room and Door Detection Failure Cases

6 Conclusion

We have presented a method for floor plan analysis with the goal of automatically extracting topological spatial information. The algorithm focuses on extracting text before flood filling to find the rooms and hallways. By using SURF features to match to a door symbol template, the connections between rooms were found. This is enough information to construct a topological and topometric graph from the floor plan. The topological graph is useful where the metric relationships are unnecessary, such as for human-robot interaction and receiving feedback from the robot on what path options are available. It may also be used in high level path planning in scenarios involving previously unexplored areas. The extracted graph is dependent on successful text detection and extraction but is robust enough to function with some failures, such as elevators or stairwells.

Future work will focus on increasing the robustness of this method, including testing on architectural floor plans with different styles, such as those from international sources. We aim to improve the door detection for edge cases such as sliding doors and overlapped doors. Other methods for room fill may also be investigated, which may handle the semantics of multiple interconnected hallways as multiple spatial regions.

Given sufficiently high quality images, our algorithm could also be applied to images captured by the robot of floor plans, signs or maps posted on walls and doors. This will allow a robot to navigate and reason about its environment without prior information.

References

[Ahmed *et al.*, 2011] Sheraz Ahmed, Marcus Liwicki, Markus Weber, and Andreas Dengel. Improved automatic analysis of architectural floor plans. In *2011 International Conference on Document Analysis and Recognition (ICDAR)*, pages 864–869. IEEE, 2011.

[Aydemir *et al.*, 2012] Alper Aydemir, Patric Jensfelt, and John Folkesson. What can we learn from 38,000 rooms? reasoning about unexplored space in indoor environments. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4675–4682. IEEE, 2012.

[Bay *et al.*, 2008] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, June 2008.

[Dosch *et al.*, 2000] Philippe Dosch, Karl Tombre, Christian Ah-Soon, and Grald Masini. A complete system for the analysis of architectural drawings. *International Journal on Document Analysis and Recognition*, 3(2):102–116, 2000.

[Epshtein *et al.*, 2010] Boris Epshtein, Eyal Ofek, and Yonatan Wexler. Detecting text in natural scenes with stroke width transform. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[Galindo *et al.*, 2005] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.A. Fernandez-Madrigal, and J. Gonzalez. Multi-hierarchical semantic maps for mobile robotics. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2278–2283, Aug 2005.

[Gansner and North, 2000] Emden R. Gansner and Stephen C. North. An open graph visualization system and its applications to software engineering. *SOFTWARE - PRACTICE AND EXPERIENCE*, 30(11):1203–1233, 2000.

[Hart *et al.*, 1968] P.E. Hart, N.J. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, July 1968.

[Ito *et al.*, 2014] S. Ito, F. Endres, M. Kuderer, G.D. Tipaldi, C. Stachniss, and W. Burgard. W-rgb-d: Floor-plan-based indoor global localization using a depth camera and wifi. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014.

[Joseph *et al.*, 2013a] S.L. Joseph, Chucai Yi, Jizhong Xiao, Yingli Tian, and Fei Yan. Visual semantic parameterization - to enhance blind user perception for indoor navigation. In *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1–6, July 2013.

[Joseph *et al.*, 2013b] S.L. Joseph, Xiaochen Zhang, I Dryanovski, Jizhong Xiao, Chucai Yi, and Yingli Tian. Semantic indoor navigation with a blind-user oriented augmented reality. In *2013 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3585–3591, Oct 2013.

[Koel, 1999] L. Koel. *Construction Print Reading*. Blueprint Reading and Drafting Series. Delmar Publishers, 1999.

[Lam *et al.*, 2014] Obadiah Lam, Feras Dayoub, Ruth Schulz, and Peter Corke. Text recognition approaches for indoor robotics: a comparison. In *2014 Australasian Conference on Robotics and Automation (ACRA)*, December 2014.

[Li *et al.*, 2014] Yao Li, Wenjing Jia, Chunhua Shen, and A. van den Hengel. Characterness: An indicator of text in the wild. *IEEE Transactions on Image Processing*, 23(4):1666–1677, April 2014.

[Liu *et al.*, 2012] Ziyuan Liu, Dong Chen, and G. von Wichert. Online semantic exploration of indoor maps. In *2012 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4361–4366, May 2012.

[Luo *et al.*, 2010] R.C. Luo, Yu-Chih Lin, and Ching-Chung Kao. Autonomous mobile robot navigation and localization based on floor plan map information and sensory fusion approach. In *2010 IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pages 121–126, Sept 2010.

[Macé *et al.*, 2010] Sébastien Macé, Hervé Locteau, Ernest Valveny, and Salvatore Tabbone. A system to detect rooms in architectural floor plan images. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, pages 167–174. ACM, 2010.

[Matas *et al.*, 2004] J Matas, O Chum, M Urban, and T Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761 – 767, 2004. British Machine Vision Computing 2002.

[Mozos *et al.*, 2007] scarMartnez Mozos, Cyrill Stachniss, Axel Rottmann, and Wolfram Burgard. Using adaboost for place labeling and topological map building. In Sebastian Thrun, Rodney Brooks, and Hugh Durrant-Whyte, editors, *Robotics Research*, volume 28 of *Springer Tracts in Advanced Robotics*, pages 453–472. Springer Berlin Heidelberg, 2007.

[Posner *et al.*, 2010] I Posner, P. Corke, and P. Newman. Using text-spotting to query the world. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3181–3186, Oct 2010.

[Schulz *et al.*, 2015] Ruth Schulz, Ben Talbot, Obadiah Lam, Feras Dayoub, Peter Corke, Ben Upcroft, and Gordon Wyeth. Robot navigation using human cues: A robot navigation system for symbolic goal-directed exploration. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015.

[Smith, 2007] Ray Smith. An overview of the tesseract ocr engine. In *2007 International Conference on Document Analysis and Recognition (ICDAR)*, volume 7, pages 629–633, 2007.

[Whiting *et al.*, 2007] Emily Whiting, Jonathan Battat, and Seth Teller. Topology of urban environments. In Andy Dong, AndrewVande Moere, and JohnS. Gero, editors, *Computer-Aided Architectural Design Futures (CAADFutures) 2007*, pages 114–128. Springer Netherlands, 2007.