

Exploiting 2D Floorplan for Building-scale Panorama RGBD Alignment

Erik Wijmans
Washington University in St. Louis
<http://cv2.wustl.edu/~erikws/>

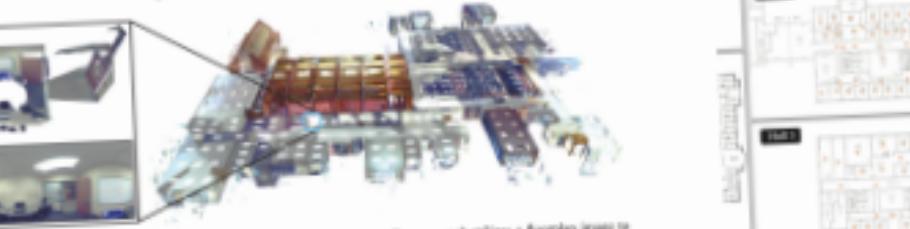


Figure 1: The paper tackles building-scale panorama RGBD image alignment. Our approach utilizes a floorplan image to significantly reduce the number of necessary scans and hence human operating costs.

Abstract

This paper presents a novel algorithm that utilizes a 2D floorplan image to align panorama RGBD scans. While effective panorama RGBD alignment techniques exist, such a system requires extremely dense RGBD image sampling. Our approach can significantly reduce the number of necessary scans with the aid of a floorplan image. We formulate a novel Multi-Feature Block Inference problem as a scene placement over the floorplan, as opposed to the conventional scan-to-scan alignment. The technical contribution lies in the multi-modal image correspondence cues (between scans and schematic floorplans) as well as a novel coverage potential avoiding an inherent scaling bias. The proposed approach has been evaluated on five challenging large-scale datasets. To the best of our knowledge, we present the first effective system that utilizes a 2D floorplan image for building-scale 3D pointcloud alignment. The source code and the data will be shared with the community in further enhance indoor mapping research.

1. Introduction

3D scanning hardware has made remarkable progress in recent years, where successful products exist in industry for commercial applications. In particular, Panorama RGBD

scanners have found real-world applications as the system produces both 3D geometry and interactive panoscopic images. For instance, Faro 3D [21] is a professional-grade panorama RGBD scanner, which can reach more than 100 meters and produce 100 million points per scan within a sufficient accuracy. The device is perfect for 3D reconstruction, documentation, or surveillance in indoor mapping, civil engineering or GIS applications. Matcaput [3, 4] is an emerging low-end solution that can reach only 5 meters, but is much quicker (i.e., 1 to 2 minutes per scan), and has demonstrated compelling results for Real Estate markets.

2. Related work

Two approaches exist for indoor 3D scanning: “RGBD scanning” or “Panorama RGBD scanning”. RGBD scanning continuously moves a depth camera and scans a room. This has been the major choice among Computer Vision researchers [16, 7, 21] after the success of Kinect Fusion [17].

The input is a RGBD video stream, where Simultaneous Localization and Mapping (SLAM) is the core technology. Panorama RGBD scanning has been rather successful in industry, because (1) data acquisition is easy (i.e., picking a 2D position as opposed to a 6 DoF navigation in RGBD scanning); (2) alignment is easier thanks to the panoramic field of view; and (3) the system provides panoscopic images, essential for many visualization applications. Structure from Motion (SfM) is the core technology in this approach. This paper provides an automated solution for Panorama RGBD alignment, and the consistency of the solution focuses on the description of the SfM techniques, where we refer the reader to a survey article [10] for the SLAM literature.

Different from standard SfM formulation, we do not know which pairs of variables (i.e., scans) should have interactions, because our variables encode the placements of the scans. Therefore, we set up a potential for every pair of scans. The potential measures the photometric and geometric

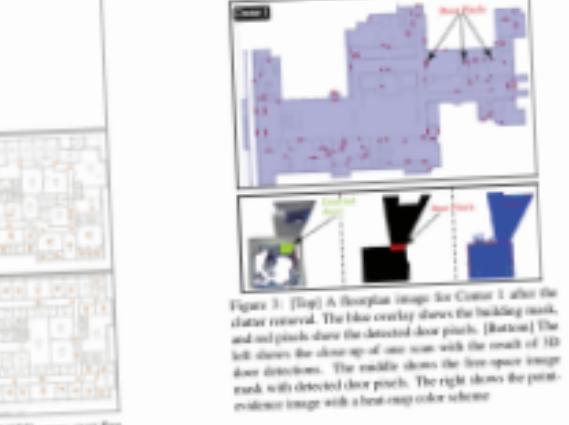


Figure 2: (Top) A floorplan image for Center 1 after the clutter removal. The blue overlay shows the building mask, and red pixels show the detected door pixels. (Bottom) The left shows the cleanup of one scan with the result of 3D door detections. The middle shows the free-space image mask with detected door pixels. The right shows the point-evidence image with a heat-map color scheme.

pixel distance inside the mask if the pixel is between the left and the right most pixels in its row and between the top and the bottom most pixels in its column.

2.8. The average penalty over all the door pixels in the evidence image is the semantic penalty.

Geometric cue: Measuring the consistency between the floorplan image and the point-evidence image is a trivial challenge: (1) A floorplan image contains extra symbols that are not in evidence images; (2) An evidence image contains obstacles that are not in a floorplan image; (3) The style of a floorplan (e.g., line thickness) may vary; and (4) Both are essentially line-drawings, making the comparison subject to small errors. In practice, we have found that the following consistency potential provides a robust metric:

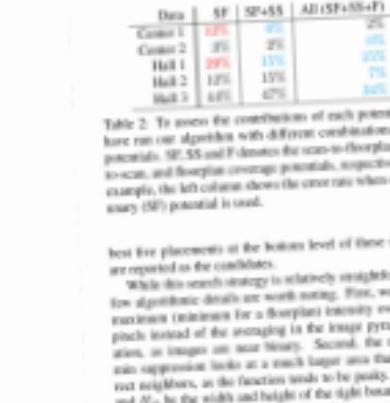
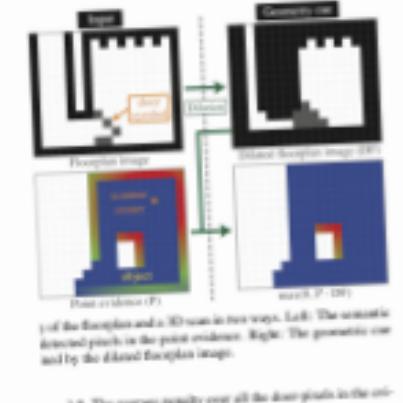
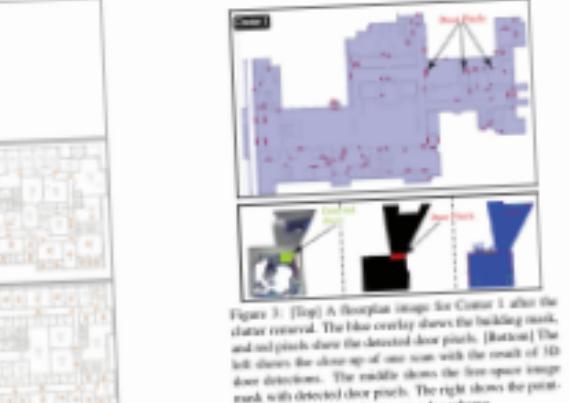
We first apply a standard morphological dilation operation (a 5x5-px kernel) to a floorplan image, using the OpenCV default implementation with a 5×5 kernel. We then measure how much of the point-evidence image is NOT explained by the floorplan, by (1) subtracting the dilated floorplan image (DF) from the point-evidence image (P); then (2) changing the regular intensities to 0. The sum of intensities in this residual image (over Ω , $P - DF$) divided by the sum of intensities in the original evidence image (P) calculates the measure of the discrepancy. We sweep the size of the floorplan and the point-evidence image, compute the size of the floorplan and the point-evidence image, and then the MRF optimization iterates until the residual scan placement is below a threshold.

2.9. The floorplan images that replace the scan-to-floorplan (semantic) truth is in the top 1 or 2 placements based on the MRF optimization with the replaced unary potential. Note SSD (sum of squared differences) metric is not useful in the placement with the best unary potential. The optimization usually converges after 50 iterations.

4. MRF formulation

The multi-modal nature of the problem makes our formulation fundamentally different from existing ones [21, 22]. The first critical difference lies in the definition of the variables. In existing approaches, a variable encodes a 3D relative placement between a pair of scans [21, 22]. In our formulation, a variable encodes a 2D absolute placement of a single scan over a floorplan image.

Let $\mathcal{S} = \{s_1, s_2, \dots\}$ be our variables, where s_i encodes the 2D placement of a single scan, s_i , consisting of two components: (1) rotation, which takes one of the four angle values corresponding to the number of the input scans. For large datasets with 30 to 80 scans, these steps roughly take 3 hours, 2.5 hours, and 30 minutes, respectively. The pre-processing in the bottleneck that is 10 to 100 times faster than the MRF optimization serves to shuffle around scan placements including correct ones to achieve a low-energy state. Nonetheless, the total potential, in particular the magnitude of the total potential divided by the number of scans is a good indicator of success. The quantity for Hall 3 is a few times larger than the others and indicates that “something is wrong”. Our main future work is to develop a robust algorithm to detect potentially erroneous scan placements, which will allow a quick user feedback to correct mistakes. We will share our source code and high-end building-scale datasets to further enhance indoor mapping research.



Data	SF	SF+SS	All (SF+SS+F)
Center 1	0.95	0.95	0.95
Center 2	0.95	0.95	0.95
Hall 1	0.95	0.95	0.95
Hall 2	1.00	1.00	0.95
Hall 3	0.95	0.95	0.95

Table 2: To assess the contribution of each potential, we have run our algorithm with different combinations of the potentials. SF and SS denotes the scan-to-floorplan, scan-to-scan, and floorplan coverage potentials, respectively. For example, the left column shows the error rate when only the unary (SF) potential is used. It is worth noting that expanding the candidates did not help in reducing the error rate for Top 3, because false cases are usually outliers.

best five placements at the bottom level of these searches are reported as the candidates.

While this search strategy is relatively straightforward, a few algorithmic details are worth noting. First, we use the maximum (minimum) for a floorplan intensity over 2×2 pixels instead of the averaging in the image pyramid creation, as images are near binary. Second, the non-local min suppression looks at a mask larger than the direct neighbors, as the function tends to be noisy. Let W_D and H_D be the width and height of the tight bounding box containing the floorplan mask. We look at a square region whose size is $(W_D + H_D)/10$. Third, the threshold at the non-local min suppression is the mean minus the standard deviation of the evaluated scores at this same pyramid level. Fourth, we speeded up the unary potential evaluation by skipping some placements when more than 10% of the corresponding free-space mask goes outside the building mask. Lastly, $T = 7$ children pixels (every other pixel is chosen at the perimeter for sparsification of 2×2 are searched under each level iteration for more robustness).

When the placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 8 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Our method is not perfect and has exposed several failure modes. First, our approach tends to make mistakes for small storage-style rooms, where a small room with a lot of clutter makes the geometric cue very noisy. Second, there are genuinely ambiguous cases where the scene geometry, appearance, and floor locations are exactly the same. Lastly, our method has made major errors (incorrectly, simply because the floorplan has not reflected constraints in the past). Unfortunately, it was difficult to identify erroneous scans based on the potentials. As the presence of problematic scans triggers the MRF optimization, it seems to shuffle around scan placements including correct ones to achieve a low-energy state. Nonetheless, the total potential, in particular the magnitude of the total potential divided by the number of scans is a good indicator of success. The quantity for Hall 3 is a few times larger than the others and indicates that “something is wrong”. Our main future work is to develop a robust algorithm to detect potentially erroneous scan placements, which will allow a quick user feedback to correct mistakes. We will share our source code and high-end building-scale datasets to further enhance indoor mapping research.

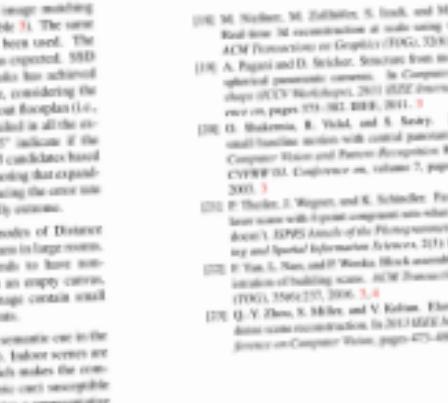
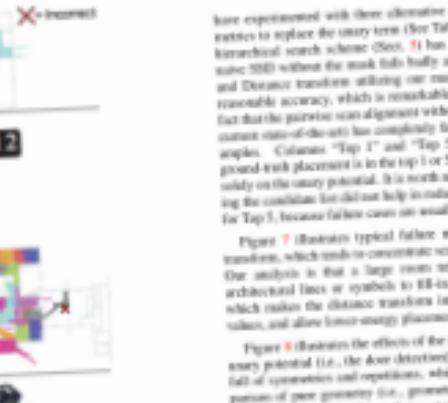
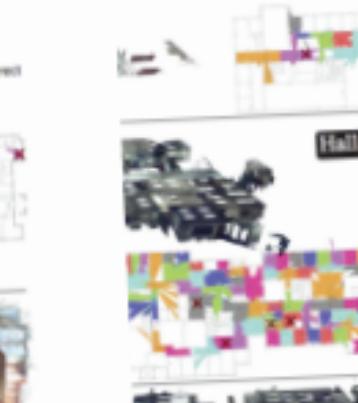
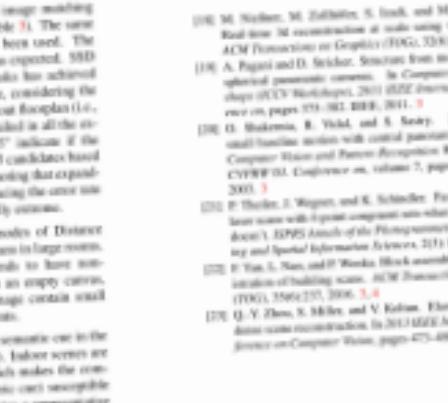
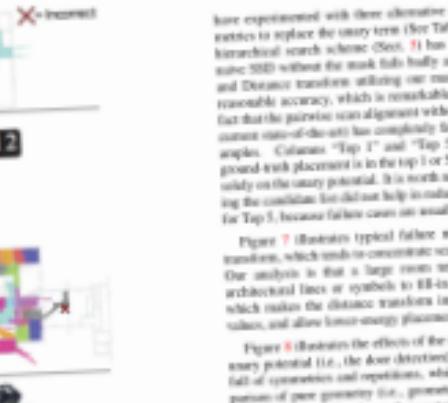
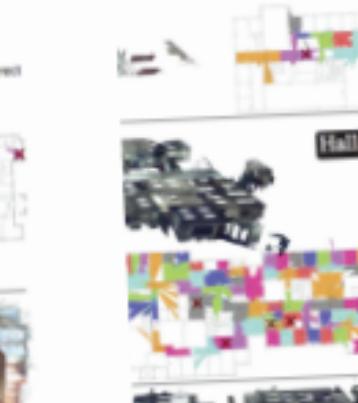


Figure 7 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 8 illustrates the effects of the semantic cue in the unary potential (i.e., the door detection). Indoor scenes are full of symmetries and repetitions, which makes the computation of pure geometry (i.e., geometric cues) susceptible to local minima. The figure demonstrates a representative case, where the door detection break such an ambiguity.

Figure 9 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 10 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 11 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 12 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 12 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 13 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 14 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 15 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 16 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 17 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 18 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 19 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 20 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 21 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 22 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 23 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 24 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 25 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 26 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 27 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 28 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 29 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 30 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 31 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 32 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 33 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 34 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 35 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 36 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 37 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 38 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 39 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 40 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 41 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 42 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 43 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 44 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 45 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 46 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 47 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 48 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 49 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 50 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 51 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 52 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 53 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 54 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 55 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 56 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 57 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 58 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 59 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 60 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 61 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 62 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 63 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 64 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 65 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 66 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 67 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 68 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 69 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 70 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 71 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 72 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 73 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 74 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 75 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 76 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 77 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 78 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 79 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 80 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 81 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 82 illustrates the effects of the semantic cue in the floorplan coverage potential. The placement is ambiguous even with the geometric and the semantic cues, we rely on the MRF optimization with the full three potentials. Figure 83 compares the final scan placements with or without the floorplan coverage potential. The floorplan coverage potential tends to avoid “stacking” and evenly distribute the placements.

Figure 84 illustrates typical failure modes of Distance transform, which tends to concentrate scans in large rooms. Our analysis is that a large room tends to have non-architectural lines or symbols to fill in an empty canvas, which makes the distance transform image contain small values, and allow lower-energy placements.

Figure 85 illustrates the effects of the semantic