

Report of Box Detection

Author: Zhiyi Tang, Yifei Li

Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

Date: October 2025

1. Results

1.1 Visualization Results

To better illustrate the detection results of floor and box planes, Figure 1 shows the visualization of the estimated planes for four different Kinect datasets. The green area corresponds to the detected floor, the red area to the box top, and the blue regions represent remaining background points.

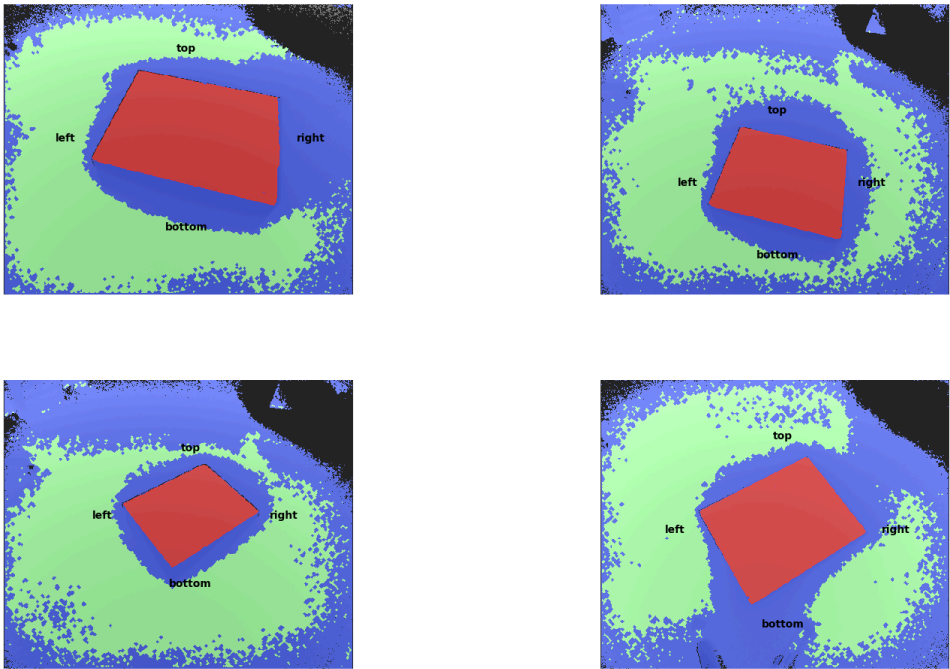


Figure 1: Visualization of detected planes (top: datasets 1–2; bottom: datasets 3–4).

1.2 Quantitative Results

Dataset	Height (m)	Length (m)	Width (m)
example1kinect	0.196	0.480	0.310
example2kinect	0.196	0.474	0.306
example3kinect	0.193	0.532	0.370
example4kinect	0.192	0.492	0.376

2. Limitations and Recognition Errors

Although the proposed RANSAC-based plane detection method works robustly for most datasets, there are still some recognition errors and unstable results, as illustrated below.

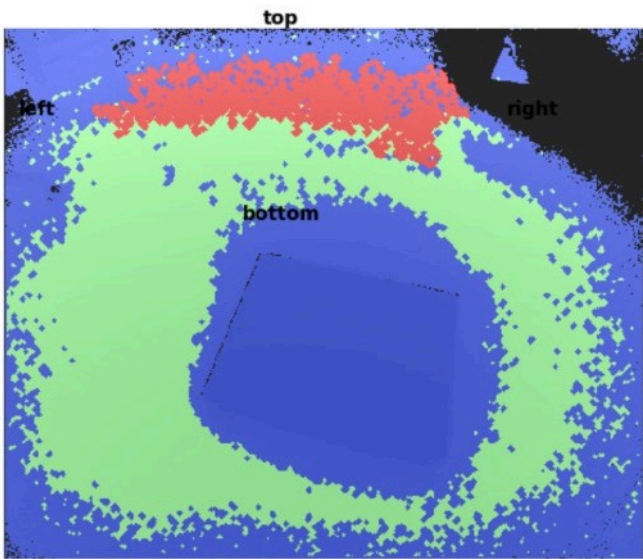


Figure 2: Example of recognition failure on `example3kinect.mat`.

2.1 Quantitative Evidence

Dataset	Height (m)	Length (m)	Width (m)	Observation
example1kinect	0.188	0.482	0.315	Correct detection
example2kinect	0.190	0.479	0.311	Correct detection
example3kinect	0.056	1.452	1.078	Wrong top plane detected
example4kinect	0.184	0.495	0.378	Correct detection

The third dataset shows a clear recognition error: the algorithm mistakenly fitted a large background plane parallel to the floor instead of the actual box top. As a result, the estimated height collapsed to only **5–6 cm**, while the length and width exploded to over **1 m**, which does not match the real box dimensions.

2.2 Possible Causes

1. **Dominant plane confusion**

In `example3kinect.mat`, a large background surface is parallel to the ground and covers more pixels than the top of the box, so RANSAC naturally selects it as the "top" plane because it contains more inliers.

In Kinect data, there may be reflections or edge points near the ground (such as small bumps on the ground). These points also meet the "parallel to the ground" requirement in non-ground areas and are therefore incorrectly identified as "top of the box."

2. Fixed global RANSAC threshold

The distance threshold (`0.005 m`) is static and may not adapt well to local noise or small box regions, which makes thin structures less competitive in the model evaluation.

2.3 Improvement Suggestions

To make the algorithm more **robust, accurate, and stable**, we propose the following improvements:

1. Filtering near-ground false positives

Since background planes parallel to the floor often have very small height differences (typically 1–4 cm), an additional constraint can be applied to exclude points whose distance to the detected floor is below a certain threshold (e.g., 5 cm). This simple rule effectively removes near-ground "false top" candidates..

2. Adaptive threshold

Use residual statistics (MAD or standard deviation of inliers) to adapt the RANSAC threshold per dataset

3. Summary

In summary, while the method performs well on most datasets, its main limitation lies in **ambiguous top-plane selection** when large background planes exist. Adding adaptive filtering and local re-fitting would significantly improve robustness without sacrificing simplicity.