S23DR 2025 Challenge: A Handcrafted 4th Place Solution

Anatolii Shokha
Lviv Polytechnic National University
anatolii.shokha.it.2021@lpnu.ua

Abstract

This report covers my handcrafted solution for the S23DR 2025 Challenge. I built on the official baseline by improving edge detection, using a better depth assignment method, and refining the 3D merging. The pipeline uses only classical computer vision, with no learned models. My final submission ranked 4th on the private leaderboard, achieving a 25% higher HSS than the baseline. I also describe several more complex strategies that I tested but which failed, suggesting that for this task, simple and robust methods worked best.

1. Introduction

The goal of the S23DR 2025 Challenge [1] was to convert multi-view images and sensor data into structured 3D wireframes. My approach was to build a simple, fully handcrafted pipeline, relying only on classical computer vision. Instead of adding architectural complexity, I focused on making targeted improvements to the official baseline in edge detection, depth assignment, and 3D merging. Careful parameter tuning was the key to outperforming the baseline.

2. Pipeline overview

The proposed solution takes as input the provided segmentation masks, depth maps, and COLMAP reconstruction.

Vertex detection. Candidate vertices are identified using color thresholding and connected components on the segmentation masks.

Edge detection. For edge detection, binary masks for each edge class are processed with morphological operations and robust line fitting. Vertices are associated with nearby edge segments to form initial 2D connections.

Depth assignment. Depth values are assigned to each vertex using a distance-weighted search in the sparse depth map, with fallback to the dense map if needed.

3D projection. The 2D vertices are then projected to 3D using camera parameters.

Multi-view merging and refinement. After processing all views, 3D vertices and edges are merged by clustering spatially close vertices of the same type. The wireframe is then refined by pruning outliers and enforcing edge length constraints.

3. Key methods and improvements

3.1. Improved edge morphology

I replaced the baseline's basic morphological operations with a sequence of closing followed by opening, which helped fill small gaps and remove noise in the edge masks. This led to more robust edge detection and improved connectivity.

3.2. Distance-weighted depth assignment

For each vertex, I assigned depth using a distance-weighted search in the sparse depth map, rather than simply taking the nearest value. This provided more stable and accurate depth estimates, especially in regions with sparse data.

3.3. Robust 3D merging and pruning

I refined the merging of 3D vertices by clustering spatially close points of the same type and pruning outliers based on distance to the COLMAP point cloud. This reduced duplicate vertices and improved the overall structure of the wireframe.

3.4. Parameter tuning

All thresholds and parameters were tuned empirically on the validation set to balance recall and precision, leading to the best overall Hybrid Structure Score.

4. Metric understanding and design choices

The S23DR challenge uses the Hybrid Structure Score (HSS), defined as the harmonic mean of vertex F1 and edge IoU, as the main evaluation metric. This metric rewards solutions that achieve both accurate vertex localization and correct edge connectivity.

In developing my pipeline, I focused on balancing recall and precision for both vertices and edges, as missing vertices or edges can significantly lower the HSS. Parameter

Strategy	HSS Change	Key Observation
Multi-model ensemble	-10%	Introduced too many false positives.
Strict cross-view consistency	-74%	Ignored valid occluded vertices.
Aggressive parameter filtering	-20% to -60%	Removed correct geometry with noise.
Complex geometric processing	-12% to -25%	Less robust to noise than simple methods.

Table 1. Summary of failed experimental strategies and their impact on the baseline Hybrid Structure Score (HSS).

tuning was guided by validation HSS, rather than optimizing for F1 or IoU alone.

In particular, I found that slightly increasing recall (even at the cost of some false positives) led to higher HSS, as the metric penalizes missing structure more than extra predictions.

5. What didn't work

Several strategies were explored but ultimately underperformed the baseline. While they didn't help, they revealed important things about the data and the limits of the pipeline. The impact of these experiments is summarized in Table 1.

5.1. Multi-model ensemble

I tried combining a high-precision model (anchored by COLMAP points) with my high-recall baseline, hoping to boost both metrics. However, this approach dropped the HSS by 10% and the Corner F1 by 20%. The main issue was that the ensemble introduced too many false positives, confirming that simply increasing the vertex count is not effective if precision suffers.

5.2. Strict cross-view consistency

To enforce geometric consistency, I experimented with requiring each vertex to be detected in at least two views. This led to a massive performance drop, with HSS falling by 74% and vertex/edge counts dropping over 70%. This strict rule failed because it doesn't account for occlusions or viewpoint changes, which cause many valid vertices to appear in only a single image.

5.3. Aggressive parameter filtering

I also tested making all filters stricter by reducing thresholds and adding harsh validation rules. This consistently lowered HSS by 20-60% across different tests. The takeaway was that the initial detections from the segmentation masks are too noisy to withstand such strict filtering; it removed correct geometry along with the noise.

5.4. Complex geometric processing

Replacing simple components with more complex geometric techniques, such as contour analysis or RANSAC plane fitting, also resulted in a 12-25% drop in HSS. I found that the simple, moment-based centroids were actually more robust, as the advanced methods were less tolerant of noise and often shifted vertices to incorrect positions.

5.5. Key takeaways

Although these strategies didn't improve performance, they revealed several important constraints of the problem:
- The noise in the provided segmentation masks was the biggest limiting factor. More complex methods often failed simply because they couldn't handle this noise.
- The HSS metric heavily penalizes missing geometry, so it was almost always better to accept a few false positives than to risk losing a true positive.
- The final score was highly sensitive to small changes in parameters like distance thresholds, which made careful tuning one of the most critical steps.

Overall, the most effective improvements came from targeted refinements and tuning, rather than from adding architectural complexity.

6. Conclusion

My handcrafted pipeline reached a Hybrid Structure Score (HSS) of 0.3309 on the S23DR 2025 private leaderboard, while the official baseline scored 0.264. This was enough for 4th place overall. The biggest gains came from tuning parameters and keeping the processing steps simple and robust. More complicated or aggressive methods didn't help, which shows that, for this challenge, straightforward approaches worked best.

References

[1] Jack Langerman, Dmytro Mishkin, and Huang Yuzhong. S23dr competition at 2nd workshop on urban scene modeling @ cvpr 2025. https://huggingface.co/usm3d, 2025. 1