# Centroid Distance Keypoint Detector for Colored Point Clouds *Supplementary Material*

Hanzhe Teng, Dimitrios Chatziparaschis, Xinyue Kan, Amit K. Roy-Chowdhury, and Konstantinos Karydis

University of California Riverside, CA 92521, USA

We introduce the parameters used in each method in Section A, and discuss the supplementary qualitative evaluation consisting of a few more randomly-picked frames in Section B. Results of repeatability for varying levels of noise intensity are presented in Section C. Finally, we discuss our observations about the USIP method in Section D.

## A Parameters for Each Method

Section 4.1 of the paper introduced the experimental setup and dataset-specific parameters. Herein we detail the parameters required by each method. (Parameters for the USIP method will be discussed in Section D in this supplementary material.)

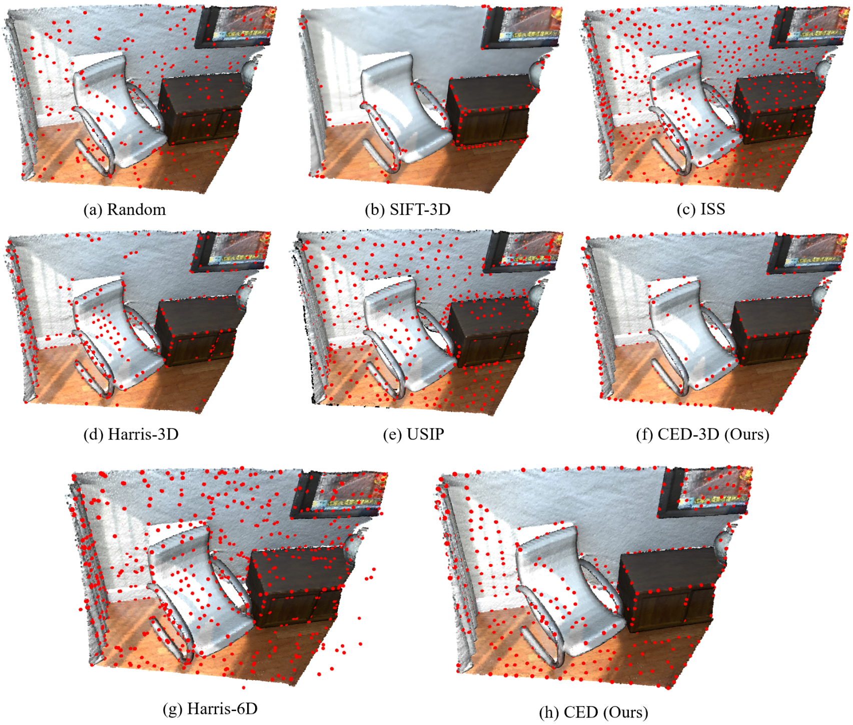
**Table 1.** Parameters for Each Method

| Method | Parameter | Value | Comment |
|---|---|---|---|
| CED/CED-3D | centroid_geo | 0.2 | Geometric centroid threshold |
| CED | centroid_rgb | 0.1 | Photometric centroid threshold |
| ISS | gamma21 | 0.975 | Ratio between 2nd and 1st eigenvalue |
| ISS | gamma32 | 0.975 | Ratio between 3rd and 2nd eigenvalue |
| Harris-3D/6D | threshold | 0.000001 | Minimum Harris response |
| SIFT-3D | min_scale | 0.01 | Minimum scale |
| SIFT-3D | n_octaves | 3 | The number of octaves |
| SIFT-3D | n_scales_per_octave | 2 | The number of scales for each octave |
| SIFT-3D | min_contrast | 0.01 | Minimum contrast |
| Random | n_points | 300 | The number of points to be picked |

A summary of parameters tuned from Redwood Synthetic dataset is presented in Table 1. They are used consistently across experiments on Redwood Synthetic [1], Redwood Scan [2] and TUM [3] datasets.

Most parameters presented herein are independent from target scenes, and the only exception is the min_scale parameter in SIFT-3D method, which is meant to match the point cloud resolution. When the scale of the environment varies significantly (e.g., in the SUN3D dataset [5]), some parameters can be adjusted to control the number of keypoints being extracted. For example, the n_scales_per_octave parameter in the SIFT-3D detector can be increased in order to obtain sufficient number of meaningful keypoints, and the centroid thresholds in CED detector can be increased to reduce the number of extracted keypoints.

The parameters used in CED, CED-3D and SIFT-3D are tuned empirically from small-scale experiments and ablation study. The two thresholds for the ISS method are set according to the recommendation in its implementation in PCL [1] and related literature (e.g., [4]). The threshold for Harris response is set following the official example in PCL. [2] The number of points picked by the random selector is set to be roughly aligned with the average number of points extracted by other methods.



(a) Random

(b) SIFT-3D

(c) ISS

(d) Harris-3D

(e) USIP

(f) CED-3D (Ours)

(g) Harris-6D

(h) CED (Ours)

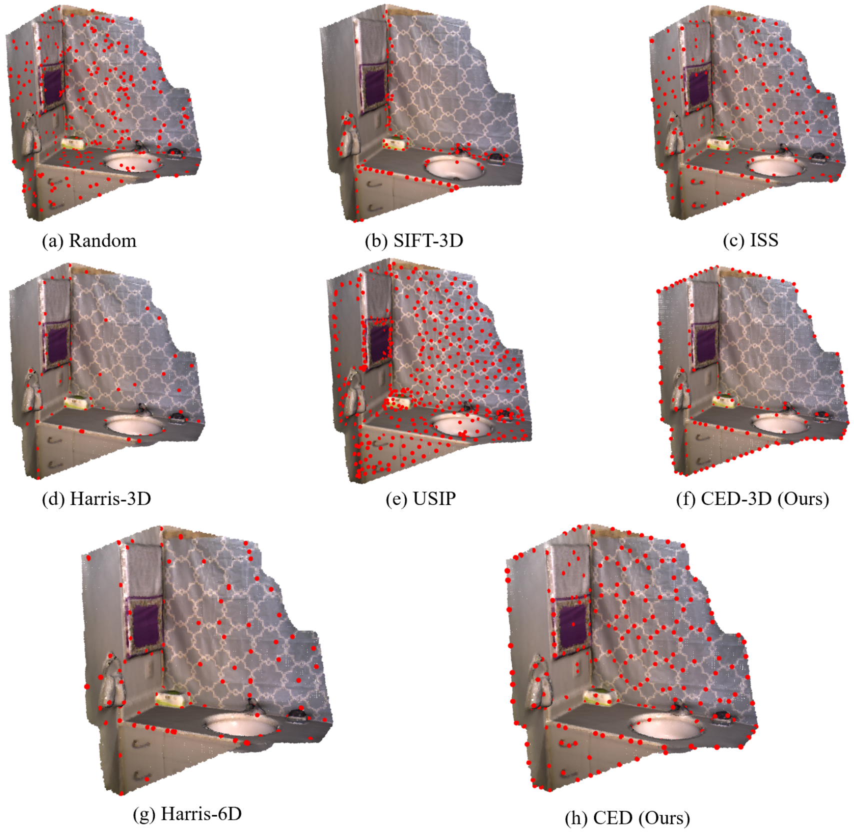
**Fig. 1.** Supplementary qualitative evaluation of an arbitrary frame of livingroom environment in the Redwood Synthetic dataset. (a-f) Methods able to extract geometry-salient keypoints only. Our proposed CED-3D keypoint detector can capture corners of the desk, the chair and the picture frame with high regularity. (g-h) Methods able to extract both geometry- and color-salient keypoints. Our proposed CED keypoint detector can further **capture shadows on the wall and the chair**, and even the photometric changes between floor tiles.

[1] Available at `https://github.com/PointCloudLibrary/pcl/blob/master/keypoints/include/pcl/keypoints/iss_3d.h#L72`

[2] Available at `https://github.com/PointCloudLibrary/pcl/blob/master/examples/keypoints/example_get_keypoints_indices.cpp#L65`

## B Supplementary Qualitative Evaluation

Section 4.2 presented the qualitative evaluation from an arbitrary frame in the Redwood Synthetic dataset. We provide herein a few more randomly-picked frames as supplementary materials, covering kitchen, bathroom and living room scenes in Redwood Synthetic and Redwood Scan datasets. Results are presented in Figure 1, Figure 2 and Figure 3. The observations made in each case are consistent with those discussed in the paper.


(a) Random

(b) SIFT-3D

(c) ISS

(d) Harris-3D

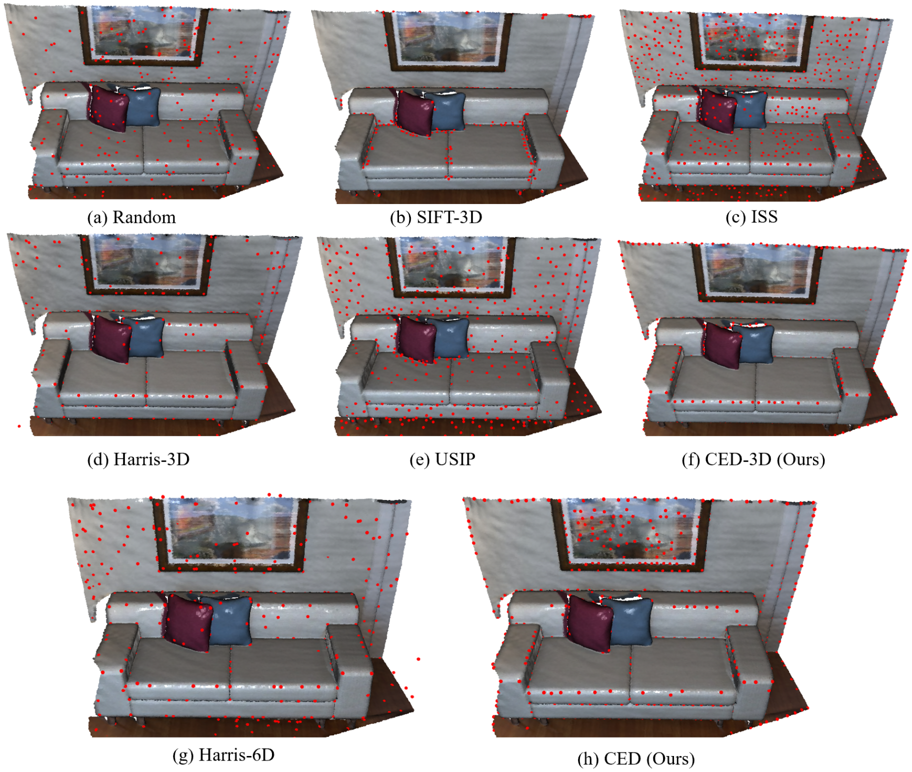
(e) USIP

(f) CED-3D (Ours)

(g) Harris-6D

(h) CED (Ours)

**Fig. 2.** Supplementary qualitative evaluation of an arbitrary frame of bathroom environment in the Redwood Scan dataset. (a-f) Methods able to extract geometry-salient keypoints only. (g-h) Methods able to extract both geometry- and color-salient keypoints. Our proposed CED detector can capture the repetitive pattern on bath towel and extract keypoints aligned with the pattern, whereas other methods extract keypoints in an uniform manner or fail to extract meaningful keypoints on the towel.


(a) Random

(b) SIFT-3D

(c) ISS

(d) Harris-3D

(e) USIP

(f) CED-3D (Ours)

(g) Harris-6D

(h) CED (Ours)

**Fig. 3.** Supplementary qualitative evaluation of an arbitrary frame of living room environment in the Redwood Synthetic dataset. (a-f) Methods able to extract geometry-salient keypoints only. (g-h) Methods able to extract both geometry- and color-salient keypoints. Our proposed CED detector can capture both corners/edges on the sofa and color changes on the picture frame, and extract keypoints from only these interested areas and leave other regions blank, whereas other methods either fail to extract keypoints on the picture frame or extract keypoints ubiquitously without clear distinction between geometry-salient, color-salient and uninterested regions.

## C Repeatability for Varying Levels of Noise Intensity

Section 4.3 discussed the evaluation of repeatability with and without noise added. In order to determine a reasonable noise level, we have conducted an experiment on Redwood Synthetic dataset with varying levels of noise intensity. The results are shown in Figure 4.

Note that this plot presents only the averaged repeatability, and does not reflect the entire statistical distribution of the repeatabilities computed from all point clouds in the dataset. (We present the results of the entire statistical distribution via the boxplot in the paper.)

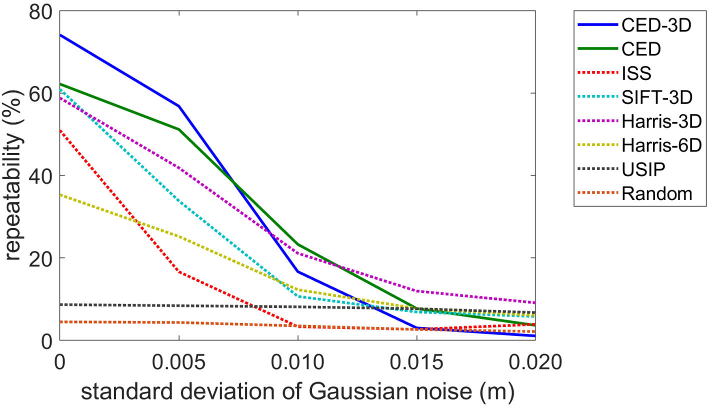


**Fig. 4.** Results of average relative repeatability evaluated on Redwood Synthetic dataset for varying levels of noise intensity. The proposed CED and CED-3D methods (solid lines) outperform other methods (dashed lines) at lower noise intensity, and the performance of all methods degrades as the noise intensity begins exceeding the point cloud resolution.

We can observe that the proposed CED and CED-3D methods (solid lines) outperform other methods (dashed lines) when the standard deviation of Gaussian noise added is less than the point cloud resolution (0.01 m for the Redwood Synthetic dataset illustrated herein). The results with standard deviation exceeding point cloud resolution are deemed less meaningful, because points in the cloud are not guaranteed to be present again within the repeatability threshold after adding noise. The boundary of most points (99.7%, according to 3-$\sigma$ rule) is at three times of the resolution, whereas the repeatability threshold is twice of the resolution. Therefore, we present in the paper only the results of repeatability with the standard deviation of Gaussian noise added to be half of the point cloud resolution.

# D Observations about the USIP Method

## D.1 Quantization Issue

As mentioned in the paper, the USIP keypoint detector works in a different way than typical keypoint detectors. Specifically, it proposes candidate positions in 3D space, as opposed to selecting existing points on the point cloud. This is partially due to constraint on quantization (i.e. approximation) in neural networks.

One direct result of this behavior is that the extracted keypoints (i.e. proposed 3D positions) can be non-deterministic given the same point cloud input. We validate this observation by performing a simple experiment, introduced as follows. Provided an arbitrary frame in the Redwood Synthetic dataset as input, we extract USIP keypoints twice and compute the repeatability, for exactly the same point cloud and without altering its pose or adding noise. The repeatability can be as low as 20% when the threshold is set to 0.02 m, and gradually increases to 100% as we relax the threshold to 0.1 m. This also explains why the repeatability of USIP can be as low as 20% on Redwood Synthetic, Redwood Scan and TUM datasets (where the repeatability $\epsilon$ is set to 0.02 m), and can increase to 50% on SUN3D dataset (where the repeatability $\epsilon$ is set to 0.2 m).

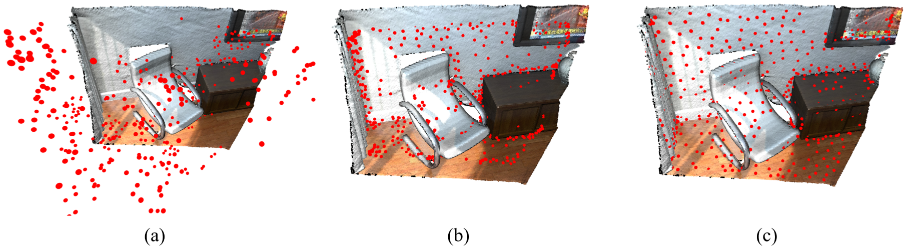
On the other hand, proposing 3D positions as keypoints can provide robustness against noise (because noise is added to existing points on the cloud), but introduce new issues for down-stream applications such as point cloud registration, where points must be selected from the original point cloud in order to estimate a meaningful transformation between two point clouds. Therefore, the points on the cloud closest to the proposed 3D positions can be selected accordingly.

This influence is minor when applied to large-scale outdoor sparse environments, but can be critical for small-scale indoor dense environments. For example, a distance offset of 10 cm between the proposed candidate 3D position and its closest point on the cloud can significantly alter the results in indoor environments, such as those in Redwood datasets, but can be negligible in outdoor scenes as in KITTI datasets.

Note that the results on Redwood Synthetic dataset reported in USIP's work are obtained by setting the repeatability $\epsilon$ to 0.1 m. It is also acknowledged by the authors (mentioned in the paper) that USIP underperforms on Redwood Synthetic dataset, and the results we obtained are consistent with their observations.

## D.2 Model Selection

We take three pre-trained models from USIP for evaluation: Oxford, 3DMatch and ModelNet. A qualitative evaluation of an arbitrary frame on Redwood Synthetic dataset is presented in Figure 5.


(a)

(b)

(c)

**Fig. 5.** Keypoints extracted by USIP method using (a) Oxford, (b) 3DMatch and (c) ModelNet pre-trained models on an arbitrary frame in the Redwood Synthetic dataset.

The results presented herein 1) indicate the lack of generalization capability of the USIP method, and 2) provide an observation that the ModelNet model fits better in our evaluated scenes.

We then evaluate the ModelNet models for point cloud registration on the Redwood Synthetic dataset using different parameters. The result is shown in Table 2. We use a notion AA-BB to denote the network structure, where AA indicates the number of keypoints produced by the neural network, and BB indicates the number of keypoints selected from AA points as output. For example, 512-256 indicates that this is a network structure that can produce 512 keypoints, and then the top 256 keypoints are selected as output according to their ordering of the saliency.

We can observe that ModelNet 512-512 performs the best, and this is the final model we adopted in our experiments across the paper.

**Table 2.** Success Rates (%) of Point Cloud Registration

| Model \ Sequence | livingroom1 | livingroom2 | office1 | office2 |
|---|---|---|---|---|
| 64-64 | 1.79 | 2.17 | 3.85 | 6.12 |
| 128-128 | 3.57 | 10.87 | 9.62 | 16.33 |
| 256-256 | 8.93 | 15.22 | 32.69 | 36.73 |
| 512-512 | **37.50** | **52.17** | **57.69** | **73.47** |
| 512-256 | 16.07 | 45.65 | 55.77 | 67.35 |
| 512-128 | 16.07 | 34.78 | 26.92 | 51.02 |

## References

1. Choi, S., Zhou, Q.Y., Koltun, V.: Robust reconstruction of indoor scenes. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 5556–5565 (2015)
2. Park, J., Zhou, Q.Y., Koltun, V.: Colored point cloud registration revisited. In: IEEE International Conference on Computer Vision. pp. 143–152 (2017)
3. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of RGB-D SLAM systems. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 573–580 (2012)
4. Tombari, F., Salti, S., Stefano, L.D.: Performance evaluation of 3D keypoint detectors. International Journal of Computer Vision **102**(1-3), 198–220 (2013)
5. Xiao, J., Owens, A., Torralba, A.: SUN3D: A Database of Big Spaces Reconstructed Using SfM and Object Labels. In: IEEE International Conference on Computer Vision. pp. 1625–1632 (2013)