Centroid Distance Keypoint Detector for Colored Point Clouds

Hanzhe Teng, Dimitrios Chatziparaschis, Xinyue Kan, Amit K. Roy-Chowdhury, Konstantinos Karydis
University of California, Riverside
United States of America
{hteng007, dchat013, xkan001, amitrc, karydis}@ucr.edu *

Abstract

Keypoint detection serves as the basis for many computer vision and robotics applications. Despite the fact that colored point clouds can be readily obtained, most existing keypoint detectors extract only geometry-salient keypoints, which can impede the overall performance of systems that intend to (or have the potential to) leverage color information. To promote advances in such systems, we propose an efficient multi-modal keypoint detector that can extract both geometry-salient and color-salient keypoints in colored point clouds. The proposed CEntroid Distance (CED) keypoint detector comprises an intuitive and effective saliency measure, the centroid distance, that can be used in both 3D space and color space, and a multi-modal non-maximum suppression algorithm that can select keypoints with high saliency in two or more modalities. The proposed saliency measure leverages directly the distribution of points in a local neighborhood and does not require normal estimation or eigenvalue decomposition. We evaluate the proposed method in terms of repeatability and computational efficiency (i.e. running time) against state-of-the-art keypoint detectors on both synthetic and real-world datasets. Results demonstrate that our proposed CED keypoint detector requires minimal computational time while attaining high repeatability. To showcase one of the potential applications of the proposed method, we further investigate the task of colored point cloud registration. Results suggest that our proposed CED detector outperforms state-of-the-art handcrafted and learning-based keypoint detectors in the evaluated scenes. The C++ implementation of the proposed method is made publicly available at `https://github.com/UCR-Robotics/CED_Detector`.

*We gratefully acknowledge the support of NSF grants #IIS-1724341 and #IIS-1901379, ONR grant #N00014-18-1-2252, and USDA/NIFA #2021-67022-33453. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

1. Introduction

Keypoint detection serves as the basis across computer vision and robotics applications such as 3D reconstruction [1,11], localization and mapping [8,12], and navigation on point clouds [9, 17]. In such applications, colored point clouds can be obtained from RGB-D cameras, where color and depth images are aligned, or camera-LiDAR systems, where they are collected after extrinsic calibration; this process is often referred to as point cloud colorization or photometric reconstruction (e.g., [21]). However, most existing keypoint detectors only consider geometric information and fail to extract color-rich keypoints, which impedes the overall performance of systems that intend to (or have the potential to) leverage color information. Hence, there is need for efficient keypoint detectors that can extract both geometric and color information in the point cloud.

Keypoint detection aims to extract a subset of points from the point cloud so that they can best represent the original data in a compact form. Some successful keypoint detectors for 2D images, such as SIFT [19] and Harris [10], have been extended to 3D space following the original design ideas by the Point Cloud Library (PCL) community [28]. However, the data structures used in 2D images and 3D point clouds are fundamentally different. This may limit deployment of such methods and has led to studies that focus on intrinsic properties of point clouds. Recent advances in deep learning has helped introduce several learning-based keypoint detectors and feature descriptors [6, 18, 37, 39]. In spite of their strong performance within their training domains, learned detectors and features may not transfer over in new scenes that are different from those used in training. For example, it may be challenging for a system trained with data collected in indoor environments to operate in diverse outdoor scenes [2].

In contrast, methods that leverage the inherent properties of point clouds may help overcome this difficulty. Several existing methods focusing on geometric properties of point clouds, such as NARF [29] and ISS [40], require eigenvalue decomposition and/or normal estimation. These operations are computationally expensive, especially when performing

keypoint detection at a large scale. In a distinct line of research, it has been shown that incorporating color modality can improve accuracy for applications such as point cloud registration [23].

The main hypothesis underlying this work is that incorporating color modality (in addition to a geometric modality) can help improve the overall performance, as the amount of information passed on to the following components in the system has increased. Despite existing descriptors that can incorporate color information (e.g., [32]), to the best of the authors' knowledge there currently exists no effective keypoint detector that can extract color-rich keypoints to feed to the descriptor. For instance, geometric-based keypoint detectors can fail to extract keypoints on a flat surface with color texture. While some methods (e.g., SIFT-3D) can extract color-rich keypoints, they do so at expense of losing geometric information (i.e. they can only respond to one modality). This limitation can be linked to lack of an effective non-maximum suppression algorithm to combine the two modalities; this is one of the key contributions of this work.

To this end, we propose an efficient multi-modal keypoint detector, named CEntroid Distance (CED) keypoint detector, that utilizes both geometric and photometric information. The proposed CED detector comprises an intuitive and effective saliency measure, the centroid distance, that can be used in both 3D space and color space, and a multi-modal non-maximum suppression algorithm that can select keypoints with high saliency in two or more modalities. The proposed saliency measure leverages directly the distribution of points in a local neighborhood and does not require normal estimation or eigenvalue decomposition. The proposed CED detector is evaluated in terms of repeatability and computational efficiency (running time) against state-of-the-art keypoint detectors on both synthetic and real-world datasets. Results demonstrate that our proposed CED keypoint detector requires minimal computational time while attaining high repeatability. In addition, to showcase one of the potential applications of the proposed method, we further investigate the task of colored point cloud registration. Results show that our CED detector outperforms state-of-the-art crafted and learning-based keypoint detectors in the evaluated scenes.

Contributions. The paper's contributions are fourfold:

- We propose an efficient multi-modal keypoint detector that can extract both geometry-salient and color-salient keypoints in a colored point cloud, with the potential to be extended and applied to point clouds with multiple modalities (e.g., colored by multi-spectrum images).
- We propose to use an intuitive and effective measure for keypoint saliency, the distance to centroid, which can leverage directly the distribution of points and does not require normal estimation or eigenvalue decomposition.
- We develop a multi-modal non-maximum suppression algorithm that can select keypoints with high saliency in two or more modalities.
- We demonstrate through experiments in four datasets that the proposed keypoint detector can outperform the state-of-the-art handcrafted and learning-based keypoint detectors.

2. Related Work

3D keypoint detectors can be categorized as those extending designs originally developed for 2D images [10,19], and those native to 3D point clouds [4, 20, 31, 40] and 3D meshes [3, 34, 38]. Following the design in 2D images, Harris family [10] computes covariance matrices of surface normal or intensity gradient in 3D space, and in 3D and color space (herein referred to as 6D space). SIFT [19] applies the difference of Gaussian operator in scale-space to find keypoints with local maximal response. However, for 3D point clouds, the amount and position of points within the spherical region are uncertain, making it hard to obtain gradients. In 3D space, Normal Aligned Radial Feature (NARF) [29] measures saliency from surface normal and distance changes between neighboring points. Intrinsic Shape Signature (ISS) [40] and KeyPoint Quality (KPQ) [20] perform eigenvalue decomposition of the scatter matrix of neighbor points and threshold on the ratio between eigenvalues. Heat Kernel Signature (HKS) [31] and Laplace-Beltrami Scale-space (LBSS) [34] measure the saliency from the response to the Laplace-Beltrami operator in the neighborhood. Local Surface Patches (LSP) [4] leverages local principal curvatures to construct the Shape Index (SI) [7] as the measure of saliency. As in SIFT, MeshDoG [38] and Salient Points (SP) [3] apply the difference-of-Gaussian operator to construct the scale space for saliency measure. We refer readers to the comprehensive evaluation in [33] for more details.

In summary, the existing methods often apply an operator to obtain point normal, curvature and gradient in the local region, and threshold on either a combination of the obtained measures or the eigenvalues of the covariance matrices. On the contrary, our proposed method leverages directly the point distribution and statistics in 3D space and color space, without the need of normal estimation or eigenvalue decomposition.

Learning-based approaches, such as USIP [18] and 3DFeat-Net [37], have also been studied. 3DFeat-Net [37] learns a 3D feature detector and descriptor for point cloud matching using weak supervision, whereas USIP [18] trains a feature proposal network with probabilistic Chamfer loss in an unsupervised manner.

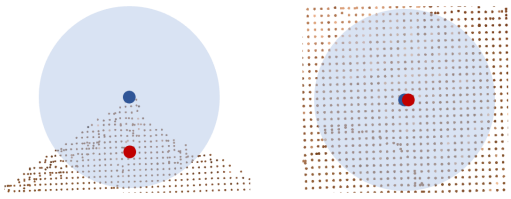
Due to the constraint on quantization in neural networks, the extracted keypoints may be non-deterministic given the same point cloud input (e.g., USIP [18]). Despite the relaxation of the requirement of ground truth, generalization capability may be challenged when deploying the system in practice. These approaches are often trained with data collected from target environments and/or designated sensors, and the performance with cross-source data in unseen scenes is unclear and less reported [13]. In addition, the high computational requirements in deep networks necessitate on-board GPUs for real-time applications, which is challenging for small robots/systems [9, 22] with limited computational resources (constrained by payload and battery life). To facilitate the deployment in various scenes and enable real-time execution, our proposed method is designed to focus on the invariant features in point clouds while keeping the computational time to a minimum.

Color modality has been taken into account in descriptors [32] as well as in registration algorithms [14, 23]. Even though color modality offers promising prospects in the aforementioned tasks, it is rarely incorporated into the design of keypoint detectors, and in turn impedes the effectiveness of methods building on top of keypoint detectors. Existing 3D keypoint detectors with color modality (e.g., Harris-6D) combine three-channel RGB responses into intensity by a weighting function; doing so leads to loss of information and hence such methods cannot capture color variations effectively. One potential reason behind this is that the commonly used non-maximum suppression algorithm can only threshold on one modality, and therefore all factors to be considered have to be weighted together into a unitary function before thresholding. In our proposed method, we consider the saliency measure in 3D space and color space separately, and then select the best response by a multi-modal non-maximum suppression algorithm that can threshold on two or more unitary functions at the same time.

3. Centroid Distance (CED) Keypoint Detector

In this section we describe our proposed CEntroid Distance (CED) keypoint detector for colored point clouds based on an efficient measure of keypoint saliency–the distance to centroid. We discuss the measure in 3D space (Section 3.1) and RGB color space (Section 3.2). Then, the two saliency measures are combined via a multi-modal non-maximum suppression algorithm (Section 3.3) to select the best local keypoint.

Consider a point $\mathbf{p} = [{}^{\mathbf{g}}\mathbf{p}, {}^{\mathbf{c}}\mathbf{p}]^{\mathrm{T}}$ that consists of a geometric component ${}^{\mathbf{g}}\mathbf{p} = \{p_x, p_y, p_z\}$, and a color component ${}^{\mathbf{c}}\mathbf{p} = \{p_r, p_g, p_b\}$, where $p_x, p_y, p_z \in \mathbb{R}$ and $p_r, p_g, p_b \in [0, 1]$. A colored point cloud is defined as $\mathbf{P} = \{\mathbf{p_1}, \mathbf{p_2}, \ldots, \mathbf{p_n}\}$. Recall that the task of keypoint detection is to extract a subset from $\mathbf{P}$ such that the extracted keypoints can potentially best represent the original data in a compact yet effective form. The typical pipeline for keypoint detection is to examine every point in $\mathbf{P}$ to obtain a measure of saliency (e.g., cornerness) by evaluating its relation with neighbors in the support (i.e. search radius). Then, a non-maximum suppression algorithm is used to disable keypoints that are not salient compared with their neighbors, and at the same time, keep those with maximum values of saliency. Following this pipeline, we introduce each component of the proposed CED keypoint detector.



(a) (b)

Figure 1. Query point (blue) and geometric centroid (red) of all points in the support for (a) a query point located at corner, and (b) a query point located in the middle of a flat surface. Points in background (brown) are extracted from real-world datasets.

3.1. Centroid Distance in 3D Space

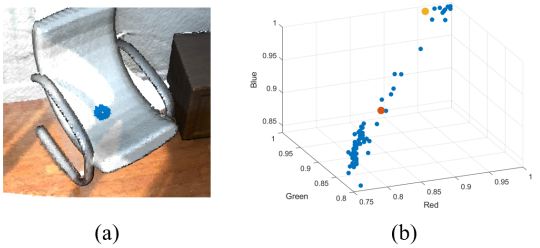
We first consider the geometric component of a point $\mathbf{p}$. Corners are among the most commonly used features in vision-based tasks. One notable measure of the orientation of a corner in 2D images is the intensity centroid [24]. In this work, we consider the centroid in 3D space, as well as color space (more details in Section 3.2). Different from those methods that take the intensity centroid as the measure of orientation, e.g., ORB [25], we observe that the distance from the origin to the centroid can be a stable measure of cornerness in 3D space.

Given a point cloud $\mathbf{P}$, and a query point $\mathbf{p_i} \in \mathbf{P}$, the set of neighbor points $\mathbf{p_j}$ within a search radius r near the query point $\mathbf{p_i}$ is defined as $N_g = \{\mathbf{p_j} \mid \|{}^{\mathbf{g}}\mathbf{p_i} - {}^{\mathbf{g}}\mathbf{p_j}\|_2 < r\}$. The geometric centroid of the spherical region ${}^{\mathbf{g}}\boldsymbol{\mu}_{\boldsymbol{p_i}}$ is

$$ {}^{\mathbf{g}}\boldsymbol{\mu}_{\boldsymbol{p_i}} = \frac{1}{|N_g|} \sum_{N_g} {}^{\mathbf{g}}\mathbf{p_j} \ , \tag{1} $$

where $|N_g|$ is the cardinality of set N_g. The geometric centroid of a spherical region is equivalent to the mean over all points in N_g in each dimension of ${}^{\mathbf{g}}\mathbf{p_j}$.

We compute the distance between the query point, $\mathbf{p_i}$, and the geometric centroid of its neighbors, ${}^{\mathbf{g}}\boldsymbol{\mu}_{\boldsymbol{p_i}}$, using the L2 norm, i.e. $d_g = \|{}^{\mathbf{g}}\mathbf{p_i} - {}^{\mathbf{g}}\boldsymbol{\mu}_{\boldsymbol{p_i}}\|_2$. Figure 1(a) demonstrates an example of a corner, where the query point (blue) has the maximum response in the vicinity, i.e. the largest distance to the geometric centroid (red). On the other hand, Figure 1(b) shows a flat surface, where the geometric centroid (red) is in close proximity to the query point (blue). It



(a) (b)

Figure 2. (a) A detected keypoint located at position with significant color variation (spherical region colored in blue). (b) RGB color space graph for neighbor points (blue) in the support of the query point (yellow) and the photometric centroid (red) according to the detected keypoint in (a).

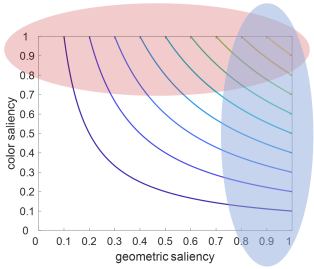


Figure 3. The contour plot for the multiplication of two saliency measures. The light blue zone reflects the geometry-salient keypoints, whereas the light red zone reflects the color-salient keypoints.

is evident that the greater this distance is, the more salient it is as a corner point.

Geometric Centroid Invariance. As the relative position (distance and angle) of all points with respect to each other within a sphere is fixed regardless of the translation and rotation of the sphere, the relative position of the geometric centroid of all points is also fixed with respect to the center of sphere, at which the query point is located. Hence, the L2 norm from the query point to its geometric centroid, used as the saliency measure in 3D space, is invariant to rotation and translation.

3.2. Centroid Distance in Color Space

Besides the geometric component, we can also consider the color component in colored point clouds (when available) by extending the concept of centroid to color space as a separate measure. Given a query point $\mathbf{p_i}$, we first identify its neighbor points $\mathbf{p_j}$ as discussed in Section 3.1. Then, we collect the color components of selected neighbor points (i.e. ${}^{\mathbf{c}}\mathbf{p_j}$, $j = 1, \ldots, |N_g|$) into a three-dimensional RGB color space set. The goal is to find the points with greatest color variation within this set. Similar to (1), we can then compute the photometric centroid ${}^{\mathbf{c}}\boldsymbol{\mu}_{\boldsymbol{p_i}}$ as

$$ {}^{\mathbf{c}}\boldsymbol{\mu}_{\boldsymbol{p_i}} = \frac{1}{|N_g|} \sum_{N_g} {}^{\mathbf{c}}\mathbf{p_j} \ . \tag{2} $$

We consider the L1 norm for the photometric centroid distance $d_c = \| {}^{\mathbf{c}}\mathbf{p_i} - {}^{\mathbf{c}}\boldsymbol{\mu}_{\boldsymbol{p_i}} \|_1$ to measure if a point has significant color changes within its neighborhood. [1]

Figure 2(a) illustrates an example of a detected keypoint with neighboring points of diverse color properties. The corresponding RGB color space graph of all points in the spherical region is depicted in Figure 2(b). Here, the photometric centroid (red) is closer to the majority of points sharing similar color. The query point (yellow) is located in the spare side, where points have larger point-to-centroid distance and are highly likely to be chosen in the following non-maximum suppression step in this neighborhood.

Photometric Centroid Invariance. Regardless of rotation and translation in 3D space, in color space the distribution of all points within the sphere remains unchanged. Hence, the L1 distance from the query point to the photometric centroid employed in color space is invariant to rotation and translation, as the measure of saliency.

3.3. Multi-modal Non-maximum Suppression

Once the saliency measures in 3D space and color space are both obtained, the next step is to determine if the query point is a salient keypoint by utilizing and integrating the two measures. The standard approach for single-response cases is to check if the current response is the maximum among its neighbors. Here, we generalize that notion to multi-modal responses. Our approach is outlined in Algorithm 1. The contour plot for the multiplication of two saliency measures is visualized in Figure 3.

Recall that the goal is to extract geometry-salient (blue ellipse in Figure 3) and/or color-salient (red ellipse in Figure 3) keypoints, which means we cannot simply remove those points that underperform in one modality, since points located at the center of a flat surface might contain rich texture and points with neighbors in same color might be located at corners. Therefore, lines 5-7 in Algorithm 1 are used to filter those points that are neither geometry-salient nor color-salient by the AND logic. After this step, points outside the red and blue ellipses, which might affect the following contour optimization, are discarded.

The next step, lines 8-13 in Algorithm 1, is to select the point that has the best performance considering both geometric and color modalities. This is achieved by optimizing the contour, constructed from the multiplication of the two saliency measures. The intuition is that in the case of a flat surface, for example, the saliency measures in 3D space are

[1] We acknowledge that a perceptually uniform space (e.g., CIELAB) may be a better choice when measuring distance in color space. However, measuring directly in RGB space is more efficient, and the performance difference is negligible when color variation is significant.

Dataset	Redwood Synthetic	Redwood Scan	TUM	SUN3D
Type	synthetic	real	real	real
Scene	RGB-D frame	stitched fragment	RGB-D frame	reconstruction
Ave. # Points	101164	63841	63211	56497
Resolution (m)	0.01	0.01	0.01	0.1
Noise added (m)	0.005	0.005	0.005	0.05
ϵ in Eq. 3 (m)	0.02	0.02	0.02	0.2

Table 1. Dataset Characteristics and Parameters

Algorithm 1 Multi-modal Non-maximum Suppression

1: **Input**: point cloud $\mathbf{P}$, saliency measure in 3D space d_{gi} and in color space d_{ci} for $\mathbf{p_i} \in \mathbf{P}$, threshold for 3D space t_g and for color space t_c
2: **Output**: keypoints $\mathbf{K}$
3: **Initialization**: $\mathbf{K} \leftarrow \{\}$
4: **for** each $\mathbf{p_i} \in \mathbf{P}$ **do**
5: **if** $d_{gi} < t_g$ **and** $d_{ci} < t_c$ **then**
6: **continue**
7: **end if**
8: maximum $\leftarrow$ *true*
9: **for** each $\mathbf{p_j} \in N_g$ **do**
10: **if** $d_{gi} \cdot d_{ci} < d_{gj} \cdot d_{cj}$ **then**
11: maximum $\leftarrow$ *false*
12: **end if**
13: **end for**
14: **if** maximum **is** *true* **then**
15: $\mathbf{K} \leftarrow \mathbf{K} \cup \{\mathbf{p_i}\}$
16: **end if**
17: **end for**

of the same magnitude, hence the points with higher color response will be selected as keypoints (multiplying with a same value does not change the order of color response). In the cases when the colors of all points are similar (e.g., table corner), the comparison becomes to mainly consider the geometric response.

Though what we illustrate in Algorithm 1 in its current form is for two modalities, the algorithm by itself is designed for multiple modalities (e.g., in the case of multi-spectrum perception). The multi-modal form can be obtained by extending line 5 and line 10 in Algorithm 1 with more saliency measures.

4. Evaluation and Results

We present the qualitative results of the proposed method against others in Section 4.2, followed by quantitative evaluation in terms of repeatability (Section 4.3) and computational efficiency (Section 4.4). To showcase one of the potential applications of the proposed CED detector, we present the results of colored point cloud registration in Section 4.5. The ablation study of the proposed method is discussed in Section 4.6.

4.1. Experiment Setup

In experiments, we consider both CED and CED-3D, which is a variant of CED that considers only geometric information for non-colored point clouds. We evaluate the performance of CED and CED-3D detectors against ISS [40], SIFT-3D [19], Harris-3D and Harris-6D [10] detectors [2] as well as a state-of-the-art learning-based keypoint detector USIP [18]. [3] A random keypoint selector is also included as baseline for comparison. Experiments are conducted with an i7 8th-gen CPU and Ubuntu 18.04 operating system. An additional Quadro P1000 GPU with CUDA 10.2 support is provided for the USIP detector only. The evaluation of computational efficiency is performed consistently using one thread.

We use four datasets, Redwood Synthetic [5], Redwood Scan [23], TUM [30] and SUN3D [35] for evaluation. The selected datasets span various scenarios including different type, scale, environment, and number of points. During pre-processing, point clouds are downsampled to their minimum resolution and NaN points are removed. Features and key parameters for each dataset are shown in Table 1. Parameters with respect to each method are tuned to their best performance or selected as recommended by the authors of the respective works. (We provide more details in the parameter tuning process for each method in the supplementary materials.)

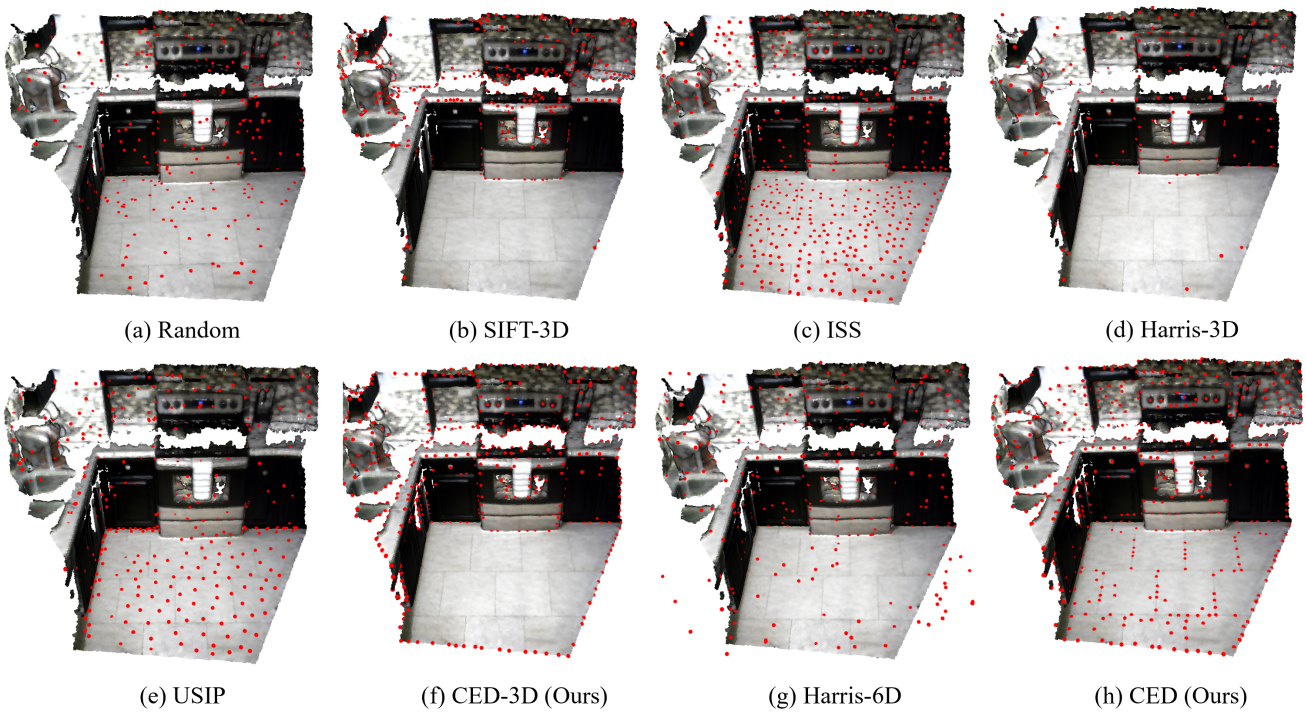
4.2. Qualitative Evaluation

We select an arbitrary frame in the Redwood Scan [23] dataset for qualitative evaluation. (The supplementary materials contain qualitative analysis for a few more randomly-picked frames across datasets.) Comparative results are shown in Figure 4. Three main observations can be made.

- The random keypoint detector produces keypoints

[2]The implementation of these methods is available in PCL 1.8, where SIFT-3D, Harris-3D and Harris-6D are implemented according to their original ideas in 2D space.

[3]The source code and pre-trained network models are provided by the authors at https://github.com/lijx10/USIP.



(a) Random (b) SIFT-3D (c) ISS (d) Harris-3D

(e) USIP (f) CED-3D (Ours) (g) Harris-6D (h) CED (Ours)

Figure 4. Qualitative evaluation of the proposed CED and CED-3D (its geometry-only variant) keypoint detectors against other methods on Redwood Scan [23] dataset. (a-f) Methods able to extract geometry-salient keypoints only. (g-h) Methods able to extract both geometry- and color-salient keypoints. Two key observations can be made in this comparison. 1) Out of all methods, only ISS, CED-3D and CED can extract keypoints on four stove knobs. 2) CED can capture **color changes between floor tiles** and extract keypoints with **high regularity**, whereas ISS and USIP extract keypoints in a somehow uniform manner, and other methods are not capable of extracting meaningful keypoints on the floor.

without using any geometric or color information in the point cloud. The USIP detector takes in non-colored point clouds and proposes candidate interest positions in 3D space (instead of selecting existing points on point clouds). The keypoints proposed by both detectors lack physical meaning and are non-deterministic given the same input cloud.

- SIFT-3D, ISS, Harris-3D and CED-3D make use of only geometric information, and can capture corners and edges in the scene. Note that only ISS and CED-3D can produce stable, regular keypoints at geometry-salient places such as the four stove knobs. However, ISS produces many meaningless keypoints on the floor due to its sensitivity to the noise on planar surfaces. As opposed to other methods, CED-3D provides meaningful keypoints with high regularity and is shown to be stable at planar surfaces.

- Harris-6D and CED can further leverage color information in the colored point cloud. Harris-6D computes color-salient keypoints using intensity gradient, but in a somehow random pattern (note that some Harris-6D keypoints are located out of the frame). In contrast, CED produces highly regular geometry- and color-salient keypoints using directly the distribution in 3D space and color space, and can clearly capture color changes between floor tiles. These regular keypoints can even serve as edge features in geometric tasks when needed. This is the key to improve the overall performance of systems that intend to leverage color information.

4.3. Repeatability

Repeatability refers to the capability to extract same keypoints under various disturbances. A stable keypoint detector is expected to be invariant to translation, rotation and noise.

We evaluate the repeatability of the proposed methods following [33]. Given a point cloud $\mathbf{P}$, we obtain point cloud $\mathbf{Q}$ by applying an arbitrary transformation matrix $T \in SE(3)$ to $\mathbf{P}$. After this step, Gaussian noise can be added to $\mathbf{Q}$ if needed in the evaluation. We then extract keypoints in $\mathbf{P}$ and $\mathbf{Q}$ to obtain keypoint sets $\mathbf{K_P}$ and $\mathbf{K_Q}$ respectively. A keypoint $\mathbf{p_i} \in \mathbf{K_P}$ is said to be *repeatable*

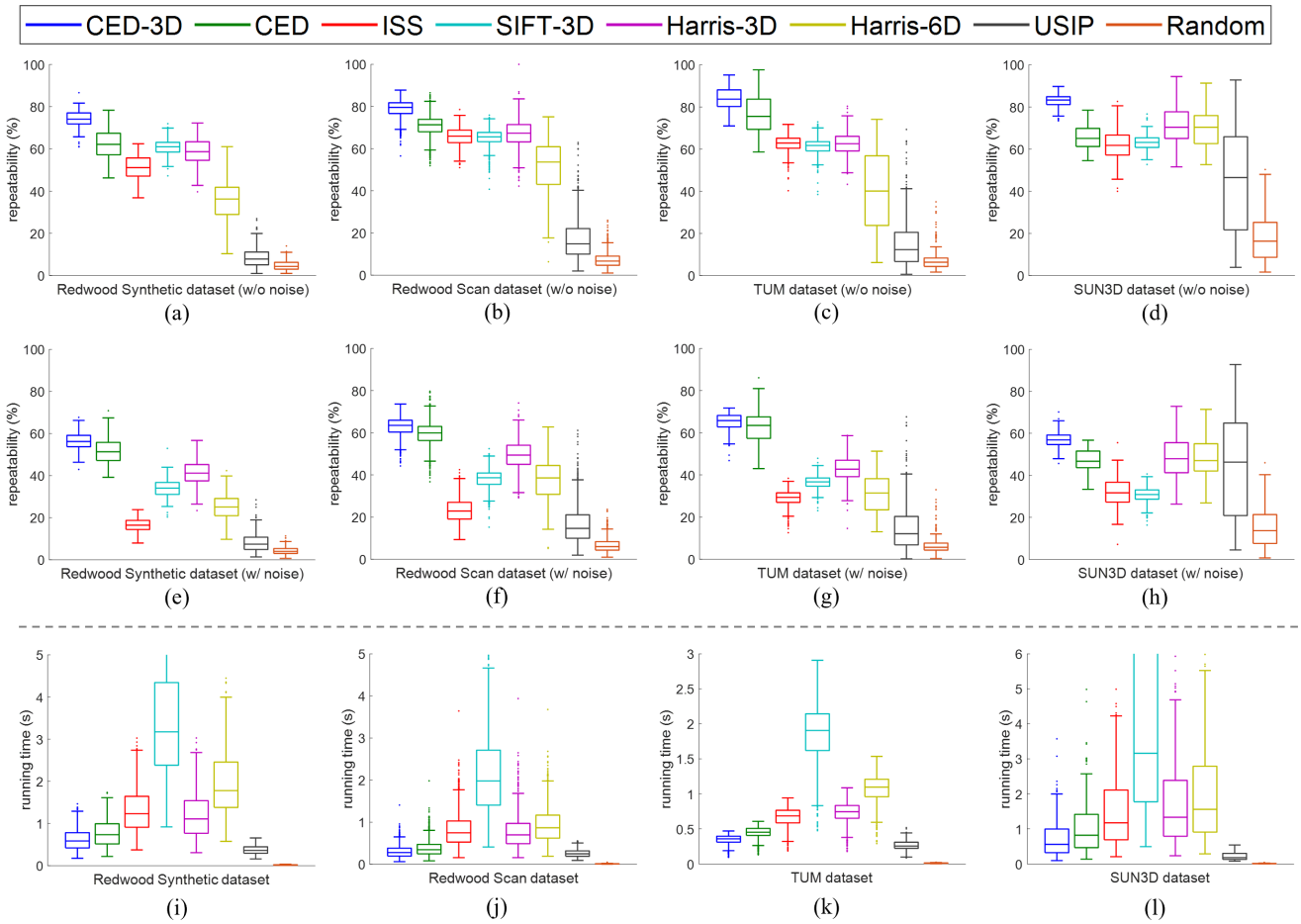


Figure 5. The evaluation result of relative repeatability without noise added (a-d), with noise added (e-h) and running time (i-l) for Redwood Synthetic [5], Redwood Scan [23], TUM [30] and SUN3D [35] datasets respectively.

if the L2 distance between the geometric components of the transformed $\mathbf{p_i}$ and its nearest neighbor $\mathbf{q_j} \in \mathbf{K_Q}$ is less than a threshold ϵ, i.e.

$$\|T\,{}^{\mathbf{g}}\mathbf{p_i} - {}^{\mathbf{g}}\mathbf{q_j}\|_2 < \epsilon \ . \tag{3}$$

The relative repeatability then is the percentage of repeatable keypoints in $\mathbf{K_P}$.

We follow the above steps to obtain the results of relative repeatability over four datasets. The standard deviation of Gaussian noise is set to be half of the resolution of point clouds, and the repeatability threshold ϵ is set to be twice of the resolution of point clouds (see Table 1 for details). As shown in Figure 5(a-h), CED and CED-3D outperform other methods by achieving the highest relative repeatability. In the case that Gaussian noise is added, our proposed methods are robust to noise, whereas significant performance degradation is observed for other methods. Additional testing for varying levels of noise intensity (up to twice the respective dataset resolution) is included in the supplementary materials, demonstrating that as the noise intensity begins exceeding the point cloud resolution, performance of all methods degrades as noise intensity increases.

4.4. Computational Efficiency

By computational efficiency we consider the running time required to extract keypoints in a point cloud. The temporal duration is recorded for each point cloud over all sequences in four datasets.

Results in Figure 5(i-l) suggest that CED and CED-3D on average keep the computation to a minimum. The random keypoint detector achieves the fastest running time (as expected); however, it has very low repeatability close to zero (shown in Figure 5(a-h)). Although the USIP keypoint detector requires similar or less time than CED and CED-3D, it requires a GPU for computation and its repeatability is significantly lower than that of CED and CED-3D.

4.5. Point Cloud Registration

The proposed CED keypoint detector can be utilized in applications such as 3D reconstruction, SLAM and odometry. To showcase one of the potential applications, we present herein results of colored point cloud registration on the Redwood Synthetic [5] dataset.

We follow a typical registration pipeline consisting of the following four steps. 1) For each pair of point clouds that describe the same scene from different view point, we apply the aforementioned eight keypoint detectors (in Sections 4.3 and 4.4) to extract keypoints. [4] 2) The feature descriptor PFHRGB [27] (the color counterpart of the popular FPFH [26] descriptor) is used to describe the keypoints. 3) Correspondences are established by nearest neighbor search in feature vector space using *k*-D tree. 4) Given the established correspondences, TEASER [36] is used to estimate a transformation that can best align the two point clouds. Registration is successful if the translation error between the estimated transformation and the ground truth is less than a threshold of 0.2 m and the rotation error is less than a threshold of 5 deg; thresholds were selected based on values used in relevant literature, e.g., [18].

Results of registration success rates for the Redwood Synthetic dataset are shown in Table 2. We observe that CED outperforms other methods in all scenes, and CED-3D achieves competitive performance by using only the geometric information in the environments. The performance improvement indicates that our CED detector is capable of capturing a few more scenes where other methods fail to extract meaningful keypoints. Note that USIP proposes interest positions in 3D space as keypoints, as oppose to selecting points on the point cloud. This behavior is non-deterministic for the same point cloud and it can severely impede the performance of point cloud registration. Reported results are the best among all trained network models. (Additional details on the USIP's performance with sub-performing network models are provided in the supplementary materials.)

Method \ Sequence	liv rm1	liv rm2	office1	office2
Random	3.57	15.22	11.54	16.33
SIFT-3D	32.14	23.91	57.69	59.18
ISS	46.43	56.52	59.62	75.51
Harris-3D	67.86	78.26	80.77	81.63
Harris-6D	71.43	78.26	84.62	79.59
USIP	37.50	52.17	57.69	73.47
CED-3D (Ours)	60.71	56.52	80.77	79.59
CED (Ours)	**76.79**	**80.43**	**86.54**	**83.67**

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

[4] For USIP, the keypoints are selected as the closest points to its proposed candidate positions in 3D space.

4.6. Ablation Study

Our CED keypoint detector depends on two thresholding parameters, one for the distance to geometric centroid and one for the distance to photometric centroid. (See t_g and t_c at the line 5 in Algorithm 1.) The geometric threshold ranges from 0 to 1, representing the ratio of the L2 distance to centroid over the radius of the spherical neighborhood region. The photometric threshold ranges from 0 to 3, representing the L1 distance variation in color space.

We evaluate on Redwood Synthetic [5] dataset for geometric threshold ranging from 0.1 to 0.5 when photometric threshold is fixed at 0.1 (Table 3), and for photometric threshold ranging from 0.1 to 0.5 when geometric threshold is fixed at 0.2 (Table 4). We observe that the number of keypoints extracted decreases as the threshold increase. The repeatability achieves the highest when $t_g = 0.2$ and $t_c = 0.5$. The running time is shown to be less dependent on the two thresholds.

t_g	0.1	0.2	0.3	0.4	0.5
# Keypoints	669.70	616.20	584.18	529.32	488.65
Rep. (%)	60.44	62.21	61.77	60.00	59.67
Runtime (s)	0.74	0.79	0.72	0.69	0.74

Table 3. Ablation Study on Geometric Centroid Threshold t_g

t_c	0.1	0.2	0.3	0.4	0.5
# Keypoint	616.20	565.67	524.31	495.34	477.46
Rep. (%)	61.94	64.30	66.05	67.58	68.37
Runtime (s)	0.70	0.74	0.68	0.71	0.66

Table 4. Ablation Study on Photometric Centroid Threshold t_c

5. Conclusion

In this work we propose the CEntroid Distance (CED) keypoint detector that can utilize both geometric and color information in colored point clouds for keypoint detection. We evaluate the proposed method against state-of-the-art handcrafted and learning-based keypoint detection methods on four synthetic and real-world datasets. Our method is demonstrated to be effective, stable and computationally efficient. We further demonstrate our method's potential to be used in applications such as colored point cloud registration.

We anticipate that the proposed CED detector can benefit systems that are capable of leveraging color information to improve performance, and that the centroid distance used herein can also serve as a stable measure to be used in other modalities. Future work also includes extension to multi-spectrum perception, and deployment of CED for autonomous robot navigation (e.g., [15, 16]) in mixed indoor and outdoor environments.

References

[1] Chris Beall, Brian J Lawrence, Viorela Ila, and Frank Dellaert. 3D reconstruction of underwater structures. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4418–4423, 2010.

[2] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, Jos Neira, Ian Reid, and John J. Leonard. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, 32(6):1309–1332, 2016.

[3] Umberto Castellani, Marco Cristani, Simone Fantoni, and Vittorio Murino. Sparse points matching by combining 3D mesh saliency with statistical descriptors. *Computer Graphics Forum*, 27(2):643–652, 2008.

[4] Hui Chen and Bir Bhanu. 3D free-form object recognition in range images using local surface patches. *Pattern Recognition Letters*, 28(10):1252–1262, 2007.

[5] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5556–5565, 2015.

[6] Haowen Deng, Tolga Birdal, and Slobodan Ilic. PPF-FoldNet: Unsupervised Learning of Rotation Invariant 3D Local Descriptors. In *European Conference on Computer Vision*, pages 602–618, 2018.

[7] Chitra Dorai and Anil K. Jain. COSMOS-A representation scheme for 3D free-form objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(10):1115–1130, 1997.

[8] Felix Endres, Jürgen Hess, Jürgen Sturm, Daniel Cremers, and Wolfram Burgard. 3D mapping with an RGB-D camera. *IEEE Transactions on Robotics*, 30(1):177–187, 2014.

[9] Fei Gao, William Wu, Wenliang Gao, and Shaojie Shen. Flying on point clouds: Online trajectory generation and autonomous navigation for quadrotors in cluttered environments. *Journal of Field Robotics*, 36(4):710–733, 2018.

[10] Christopher G Harris and Mike Stephens. A Combined Corner and Edge Detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.

[11] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments. *The International Journal of Robotics Research*, 31(5):647–663, 2012.

[12] Albert S. Huang, Abraham Bachrach, Peter Henry, Michael Krainin, Daniel Maturana, Dieter Fox, and Nicholas Roy. Visual odometry and mapping for autonomous flight using an RGB-D camera. *Robotics Research*, pages 235–252, 2016.

[13] Xiaoshui Huang, Guofeng Mei, Jian Zhang, and Rana Abbas. A comprehensive survey on point cloud registration. *arXiv:2103.02690 [cs.CV]*, 2021.

[14] Benjamin Huhle, Martin Magnusson, Wolfgang Strasser, and Achim J Lilienthal. Registration of colored 3D point clouds with a kernel-based extension to the normal distributions transform. In *IEEE International Conference on Robotics and Automation*, pages 4025–4030, 2008.

[15] Xinyue Kan, Hanzhe Teng, and Konstantinos Karydis. Online exploration and coverage planning in unknown obstacle-cluttered environments. *IEEE Robotics and Automation Letters*, 5(4):5969–5976, 2020.

[16] Xinyue Kan, Thomas C Thayer, Stefano Carpin, and Konstantinos Karydis. Task planning on stochastic aisle graphs for precision agriculture. *IEEE Robotics and Automation Letters*, 6(2):3287–3294, 2021.

[17] Philipp Krüsi, Paul Furgale, Michael Bosse, and Roland Siegwart. Driving on point clouds: Motion planning, trajectory optimization, and terrain assessment in generic nonplanar environments. *Journal of Field Robotics*, 34(5):940–984, 2017.

[18] Jiaxin Li and Gim Hee Lee. USIP: Unsupervised stable interest point detection from 3D point clouds. In *IEEE/CVF International Conference on Computer Vision*, pages 361–370, 2019.

[19] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[20] Ajmal Mian, Mohammed Bennamoun, and Robyn Owens. On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes. *International Journal of Computer Vision*, 89(2):348–361, 2010.

[21] Faraz M Mirzaei, Dimitrios G Kottas, and Stergios I Roumeliotis. 3D LIDAR–camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization. *The International Journal of Robotics Research*, 31(4):452–467, 2012.

[22] Kartik Mohta, Michael Watterson, Yash Mulgaonkar, Sikang Liu, Chao Qu, Anurag Makineni, Kelsey Saulnier, Ke Sun, Alex Zhu, Jeffrey Delmerico, Konstantinos Karydis, Nikolay Atanasov, Giuseppe Loianno, Davide Scaramuzza, Kostas Daniilidis, Camillo Jose Taylor, and Vijay Kumar. Fast, autonomous flight in GPS-denied and cluttered environments. *Journal of Field Robotics*, 35(1):101–120, 2017.

[23] Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Colored point cloud registration revisited. In *IEEE International Conference on Computer Vision*, pages 143–152, 2017.

[24] Paul L. Rosin. Measuring corner properties. *Computer Vision and Image Understanding*, 73(2):291–307, 1999.

[25] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. ORB: An efficient alternative to SIFT or SURF. In *IEEE International Conference on Computer Vision*, pages 2564–2571, 2011.

[26] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast Point Feature Histograms (FPFH) for 3D registration. In *IEEE International Conference on Robotics and Automation*, pages 3212–3217, 2009.

[27] Radu Bogdan Rusu, Nico Blodow, Zoltan Csaba Marton, and Michael Beetz. Aligning point cloud views using persistent feature histograms. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3384–3391, 2008.

[28] Radu Bogdan Rusu and Steve Cousins. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation*, pages 1–4, 2011.

[29] Bastian Steder, Radu Bogdan Rusu, Kurt Konolige, and Wolfram Burgard. Point feature extraction on 3D range scans

taking into account object boundaries. In *IEEE International Conference on Robotics and Automation*, pages 2601–2608, 2011.

[30] Jürgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers. A benchmark for the evaluation of RGB-D SLAM systems. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 573–580, 2012.

[31] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Computer Graphics Forum*, 28(5):1383–1392, 2009.

[32] Federico Tombari, Samuele Salti, and Luigi Di Stefano. A combined texture-shape descriptor for enhanced 3D feature matching. In *IEEE International Conference on Image Processing*, pages 809–812, 2011.

[33] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Performance evaluation of 3D keypoint detectors. *International Journal of Computer Vision*, 102(1-3):198–220, 2013.

[34] Ranjith Unnikrishnan and Martial Hebert. Multi-scale interest regions from unorganized point clouds. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8, 2008.

[35] Jianxiong Xiao, Andrew Owens, and Antonio Torralba. SUN3D: A Database of Big Spaces Reconstructed Using SfM and Object Labels. In *IEEE International Conference on Computer Vision*, pages 1625–1632, 2013.

[36] Heng Yang, Jingnan Shi, and Luca Carlone. TEASER: Fast and Certifiable Point Cloud Registration. *IEEE Transactions on Robotics*, 37(2):314–333, 2020.

[37] Zi Jian Yew and Gim Hee Lee. 3DFeat-Net: Weakly Supervised Local 3D Features for Point Cloud Registration. In *European Conference on Computer Vision*, pages 607–623, 2018.

[38] Andrei Zaharescu, Edmond Boyer, Kiran Varanasi, and Radu Horaud. Surface feature detection and description with applications to mesh matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 373–380, 2009.

[39] Andy Zeng, Shuran Song, Matthias Niessner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3DMatch: Learning Local Geometric Descriptors From RGB-D Reconstructions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1802–1811, 2017.

[40] Yu Zhong. Intrinsic shape signatures: A shape descriptor for 3D object recognition. In *IEEE International Conference on Computer Vision Workshops*, pages 689–696, 2009.