Performance Evaluation of 3D Local Feature Descriptors

Yulan Guo *†, Mohammed Bennamoun †, Ferdous Sohel †, Min Lu *, Jianwei Wan *, Jun Zhang *

* College of Electronic Science and Engineering
National University of Defense Technology
† School of Computer Science and Software Engineering
The University of Western Australia

Abstract. A number of 3D local feature descriptors have been proposed in literature. It is however, unclear which descriptors are more appropriate for a particular application. This paper compares nine popular local descriptors in the context of 3D shape retrieval, 3D object recognition, and 3D modeling. We first evaluate these descriptors on six popular datasets in terms of descriptiveness. We then test their robustness with respect to support radius, Gaussian noise, shot noise, varying mesh resolution, image boundary, and keypoint localization errors. Our extensive tests show that Tri-Spin-Images (TriSI) has the best overall performance across all datasets. Unique Shape Context (USC), Rotational Projection Statistics (RoPS), 3D Shape Context (3DSC), and Signature of Histograms of OrienTations (SHOT) also achieved overall acceptable results.

1 Introduction

Local features have proven to be very successful in many vision tasks such as 3D object recognition, 3D modeling, 3D shape retrieval, and 3D biometrics [1–6]. Local features have been extensively investigated during the last few decades with the aim to design descriptors which are distinctive, robust to occlusions and clutter [7]. A local feature based algorithm typically involves two major phases: keypoint detection and feature description [8, 9]. In the keypoint detection phase, keypoints with rich information contents are first identified and their associated scales (spatial extents) are then determined [9]. In the feature description phase, local geometric information around a keypoint is extracted and stored in a high-dimensional vector (i.e., feature descriptor) [10]. Finally, the feature descriptors of one pointcloud (or range image and mesh) are matched against the feature descriptors of other pointclouds of interest to yield point-to-point feature correspondences [9].

A wide variety of 3D keypoint detectors and feature descriptors have been proposed in the literature [11, 9, 12, 8]. It is widely agreed that the evaluation of feature detectors and descriptors is very important [13]. Several 3D keypoint detector evaluations can be found in the literature, e.g., [11, 9]. Descriptiveness

and robustness have been considered as two important requirements for a qualified 3D feature descriptor (see more in Section 3.3) [14, 15]. A feature descriptor is descriptive if it is capable of encapsulating the predominant information of the underlying surface. That is, it should provide sufficient descriptive richness to distinguish one local surface from another. A feature is robust if it is insensitive to a number of nuisances which can affect the data, e.g., noise and variations in the mesh resolution [9].

Although a large number of feature descriptors have been proposed, they were originally designed for various specific application scenarios and only tested on respective datasets. It is therefore very challenging for users to select the most appropriate and application independent descriptor. Beyond performance evaluations of 2D keypoint detectors [13, 16–18], 2D local descriptors [7, 19, 17, 18], and 3D keypoint detectors [11, 20, 21, 9], several evaluations on 3D local feature descriptors can also be found in the literature. Bronstein et al. [11] and Boyer et al. [20] respectively proposed an experimental evaluation of three and four 3D local feature descriptors in the context of shape retrieval. Alexandre[22] evaluated both local and global feature descriptors on a clutter-free dataset for 3D object and category recognition. Kim and Hilton [23] presented an evaluation of 3D local feature descriptors for multi-modal data registration. Other related work include [14] and [24]. However, many of these evaluations tested only few 3D local feature descriptors and for a particular application domain. Besides, the robustness of the feature descriptors is ignored in most (if not all) of these papers.

In this paper, we present a comprehensive comparison of the state-of-the-art 3D local feature descriptors and extensively test their performance on six popular datasets. Our comparison is grounded on an established methodology which was previously adopted in the evaluation of 2D local feature descriptors in [7]. Our datasets contain a large variety of scene types acquired with different imaging techniques. The performance of these descriptors on these different datasets is analyzed and discussed. We also evaluate these descriptors in three different application contexts (namely, 3D shape retrieval, 3D modeling, and 3D object recognition). Moreover, we test the robustness of these descriptors with respect to a set of nuisances including support radius, Gaussian noise, shot noise, varying mesh resolutions, image boundary, and keypoint localization error (Section 3.3). The paper is different from the literature in several aspects. First, compared to [11, 14, 20, 22, 23], our paper includes more local feature descriptors and evaluates their performance for various applications. Second, compared to [24], our paper tested 6 additional feature descriptors and analyzed the robustness of local feature descriptors. Third, as opposed to [14, 22–24], our paper compares the performance of each local feature descriptor based on criteria which only measure the performance of the feature matching of the descriptor, irrespective of any other parts of a pipeline in a specific context. Our evaluation therefore produces a performance measure for the descriptor itself rather than the whole pipeline (e.g., recognition accuracy), as commonly used in the evaluation for 2D feature descriptors (e.g., in [19, 7, 17, 18]).

The rest of this paper is organized as follows. Section 2 presents the state-of-the-art of the 3D local feature descriptors. Section 3 describes the datasets, our evaluation criteria, and the implementation details of the tested descriptors. Section 4 presents our experimental results and analysis. Section 5 concludes this paper.

2 3D Local Feature Descriptors

A number of 3D local surface descriptors have been proposed in the literature [11, 14, 20, 22, 23]. Many algorithms use histograms to represent different characteristics of the local surface. Specifically, they describe the local surface by accumulating geometric or topological measurements (e.g., point numbers) into histograms according to a specific domain (e.g., point coordinates, geometric attributes). We therefore, categorize these algorithms into spatial distribution histogram based and geometric attribute histogram based descriptors. For a comprehensive review of the existing 3D local feature descriptors, the reader is referred to [8].

2.1 Spatial Distribution Histogram based Descriptors

Spin Image (SI) [25] The SI algorithm represents each neighboring point in the support region with two parameters α and β. The radial coordinate α is defined as the perpendicular distance to the line through the surface normal, the elevation coordinate β is defined as the signed perpendicular distance to the tangent plane of the keypoint. The $\alpha - \beta$ space is then discretized into a 2D array accumulator. Finally, the SI descriptor is generated by accumulating the neighboring points into each bin of the 2D array.

3D Shape Context (3DSC) [26] The 3DSC algorithm places a 3D spherical grid at the keypoint, with the north pole of the grid being aligned with the surface normal of the keypoint. The support region is then divided into several bins along the radial, azimuth, and elevation dimensions. The 3DSC descriptor is generated by counting up the weighted number of points falling into each bin of the grid.

Unique Shape Context (USC) [27] It is an extension of 3DSC with the goal to avoid the computation of multiple descriptors at a given keypoint . First, a Local Reference Frame (LRF) is constructed for each keypoint. Next, the local surface is aligned with the LRF in order to achieve invariance to rigid transformations. Finally, the USC descriptor is generated using the same approach as 3DSC.

Rotational Projection Statistics (RoPS) [28, 10] The algorithm first aligns the local surface with its LRF. The neighboring points on the local surface are then respectively rotated around the three coordinate axes. For each rotation, the neighboring points are projected onto the three coordinate planes to generate three distribution matrices. Each distribution matrix is further encoded with five

statistics. Finally, the RoPS descriptor is generated by concatenating all these statistics from all rotations and projections.

Tri-Spin-Images (TriSI) [29] It uses the same technique as in [10] to align the local surface with its LRF. Next, a spin image is generated using the x axis as its reference axis followed by the same procedure as the SI [25]. Then, another two spin images are also generated using the y and z axes as their reference axes. The three spin images are concatenated to form the TriSI descriptor.

2.2 Geometric Attribute Histogram based Descriptors

THRIFT [30] It is a 1D histogram over the deviation angles between the surface normal at the keypoint and the normals of the neighboring points. The contribution of each neighboring point to a particular bin of the histogram is determined by two factors: 1) the density of the point samples, and 2) the distance from the neighboring point to to the keypoint.

Point Feature Histograms (PFH) [31] It is a multi-dimensional histogram over several features of the neighboring point pairs. For each pair of neighboring points, four features are calculated using the Darboux frame and the surface normals. PFH is generated by accumulating the neighboring points in particular bins along the dimensions of the aforementioned four features. In their later work [32], one feature (i.e., distance) is excluded from the histogram of PFH to improve its robustness to the density variation of points.

Fast Point Feature Histograms (FPFH) [32] A Simplified Point Feature Histogram (SPFH) is first formulated for each neighboring point by encoding the relationships between itself and its neighboring points. The FPFH descriptor is then generated as the weighted sum of the SPFH of the keypoint and the SPFHs of the neighboring points.

Signature of Histograms of OrienTations (SHOT) [33, 34] The SHOT algorithm first aligns the neighboring points of a keypoint with its LRF. Then, the support region is divided into several volumes along the radial, azimuth, and elevation axes. For each volume, a local histogram is generated by accumulating point counts into bins according to the angles between the normals at the neighboring points within the volume and the normal at the keypoint. Finally, the SHOT descriptor is generated by concatenating all local histograms.

2.3 Other Methods

Other descriptors include 3D Tensor [1], Variable-Dimensional Local Shape Descriptors (VD-LSD) [35], 2.5D SIFT descriptor [36], SI-SIFT descriptor [37], Exponential Map (EM) [38], and Integral Invariants [39, 40]. However, 3D Tensor is defined at the center of two points rather than any point of the input mesh, it is difficult to generate 3D Tensor descriptors at a set of given keypoints. A complicated training stage is required for VD-LSD to select invariant properties. Furthermore, both 2.5D SIFT, SI-SIFT, and EM descriptors can only work on depth images with a lattice structure.

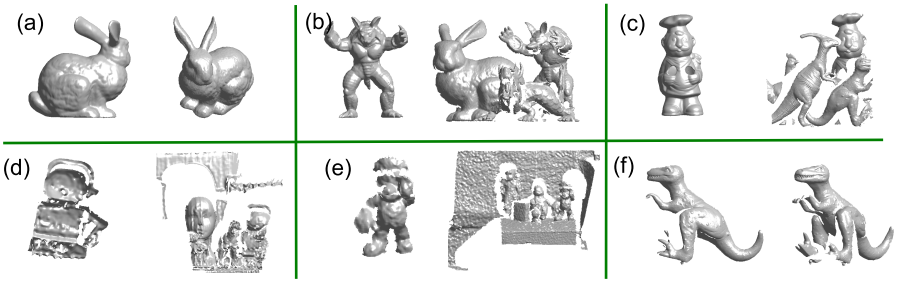


Fig. 1: Examples of models and scenes from the datasets. One model and one scene are shown for each dataset. (a) *Retrieval*. (b) *Random Views*. (c) *Laser Scanner*. (d) *Space Time*. (e) *Kinect*. (f) *2.5D Views*.

3 Experimental Setup

In this section, we describe the datasets and the evaluation criteria used in our tests to assess the performance of our selected descriptors in Section 2. We also present the implementation details of the evaluated descriptors.

3.1 Datasets

We evaluate the descriptors of Section 2 on six popular and publicly available datasets. Fig. 1 shows some examples of models and scenes taken from these datasets. The first five datasets are the same as the ones used for the 3D keypoints evaluation in [9]. Therefore, our paper provides the possibility to select an appropriate combination of 3D keypoint detectors and feature descriptors for a particular application based on their respective performance on the same dataset. We also test the descriptors on an additional dataset, i.e., the *2.5D Views* dataset.

The details of these datasets are listed in Table 1. These datasets are selected based on three major considerations: diverse acquisition techniques (e.g., Minolta Vivid 910, SpaceTime Stereo, and Kinect), different application scenarios (e.g., 3D object recognition, 3D shape retrieval, and 3D modeling), and various image qualities.

3.2 Ground-Truth

All datasets except *2.5D Views* consist of a number of models and scenes. The ground-truth rigid transformations (i.e., rotation and translation) between each model and its instance in the scene is known a priori. For more details on the generation of these ground-truth transformations, the reader is referred to [1, 33, 34, 9]. *2.5D Views* [1] contains only a set of 2.5D scenes from four objects for the reconstruction of 3D objects. The ground-truth transformation between any pair of pointclouds of the same object is first calculated by manual alignment and then refined using the iterative closest point algorithm [41].

Table 1: Datasets used in the evaluation. '-' means not relevant to that dataset.

No.	Dataset Name	Acquisition	Quality	Occlusion	Clutter	Model	Scene	# Models	# Scenes	Scenario
1	*Retrieval*[9]	Synthetic	High	No	No	3D	3D	6	18	Retrieval
2	*Random Views* [9]	Synthetic	High	Yes	Yes	3D	2.5D	6	36	Recognition
3	*Laser Scanner*[1]	Vivid	High	Yes	Yes	3D	2.5D	5	10	Recognition
4	*Space Time* [9]	SpaceTime	Medium	Yes	Yes	2.5D	2.5D	6	12	Recognition
5	*Kinect* [9]	Kinect	Low	Yes	Yes	2.5D	2.5D	27	17	Recognition
6	*2.5D Views*[1]	Vivid	High	Yes	No	-	2.5D	-	75	Modeling

3.3 Evaluation Criteria

We tested our selected descriptors (Section 2) in terms of both descriptiveness and robustness.

Descriptiveness We use the *Recall* versus *1-Precision* Curve (RPC) to evaluate the descriptiveness of a feature descriptor. RPC is commonly used in the literature for the evaluation of local feature descriptors (in both 2D images and 3D pointclouds), for example in [7, 30, 33, 34, 10]. The process for generating a RPC is described in [7]. In this paper, the Euclidean distance is used to measure the similarity between feature descriptors (as in [33, 34, 10]). Then, the nearest neighbor distance ratio based matching strategy is adopted to generate the matching features (as in [7, 10]).

In order to avoid the influence of keypoint detectors on the evaluation results, N_f keypoints are first randomly selected from each scene without keypoint detection (N_f=1000 in this paper), their corresponding model keypoints are then determined and different surface descriptors are finally generated from these fixed keypoints (as in [33, 34, 10]). Since the same procedure is applied to all methods, we believe the comparison is fair and unbiased. For *2.5D Views*, we only consider the pointcloud pairs which have an overlap of more than 50%.

Robustness We test the robustness of each feature descriptor with respect to the following variations.

Support Radius ρ: We use different support radii to define the neighboring local surface of each keypoint. For a given radius ρ, points which are distant from the keypoint by less than ρ constitute the neighboring points of that keypoint. It should be noted that in the case of 3D data, "scale" corresponds to the "support radius" [9].

Gaussian Noise: We add Gaussian noise with standard deviations of 0.1mr, 0.2mr, 0.3mr, 0.4mr, and 0.5mr to each scene, where 'mr' denotes the average mesh resolution of the models. For a given standard deviation, Gaussian noise is independently added to the x, y, and z axes of each scene point, as in [10].

Fig. 2: A mesh with its boundary shown in red.

Shot Noise: We add shot noise with outlier ratios of 0.2%, 0.5%, 1.0%, 2.0%, and 5.0% to each scene. Given an outlier ratio γ, a ratio γ of the total points in each scene are first selected and a displacement with an amplitude of 20mr is then added to each selected point along its normal direction, as in [42, 10]. Note that, shot noise usually exist in pointclouds acquired with low resolution scanners. It might be caused by miscalibration of the scanning device or image-based reconstruction of texture-less surfaces.

Varying Mesh Resolutions: We resample each scene to five levels such that only $1/2$, $1/4$, $1/8$, $1/16$, and $1/32$ of their original points are left in the resampled scene, as in [10].

Distance to the Image Boundary: We classify the scene keypoints into six groups according to their distances to the image boundary (as shown in Fig. 2). Each group contains keypoints which are within a range of distances. For example, the 2nd group contains keypoints with distances larger than $1\rho/5$ and less than $2\rho/5$ (ρ is the support radius).

Keypoint Localization Error: For each pair of corresponding points $\left(\boldsymbol{p}_i^M, \boldsymbol{p}_i^S\right)$ in each scene-model pair, we randomly select another scene point $\boldsymbol{p}_{i'}^S$ such that the distance between $\boldsymbol{p}_i^S$ and $\boldsymbol{p}_{i'}^S$ is less than a threshold τ_d. We use these new corresponding points $\left(\boldsymbol{p}_i^M, \boldsymbol{p}_{i'}^S\right)$ to produce the RPC results. Six different distance thresholds τ_d (i.e., 1mr, 3mr, 5mr, 7mr, 9mr, and 11mr) are used in Section 4.2.

3.4 Implementation Details

We use 9 different descriptors for our performance evaluation. These descriptors are briefly described in Section 2 and include SI, 3DSC, THRIFT, PFH, FPFH, SHOT, USC, RoPS, and TriSI. Some other methods presented in Section 2 have specific requirements which make their inclusion in this comparison infeasible. 3DSC, PFH, FPFH, SHOT, and USC were implemented in C++ and they are available in the Point Cloud Library (PCL) [43], while the others were implemented in Matlab (as they are not available in PCL). Note that, although SI is available in PCL, its dimensionality is fixed to 153 and different from the original paper. We therefore implemented SI in Matlab using the same parameters

(i.e., dimensionality of 225) as the original paper. In a similar manner to [16, 9, 24], the proposed default parameters in the original articles or PCL implementations were used for all selected descriptors. Unless stated otherwise in our experiments, the values of all the parameters of each descriptor were fixed when tested across all datasets. The support radius for all descriptors was set to 15mr throughout this paper (except in Section 4.2 where the "Support Radius" was varied to assess the robustness of the selected descriptors). The surface normals of the points were calculated using the method described in [44], the directions of the normals in each scene are checked to ensure that they are the same as the normals of the models. Besides, the curvatures were estimated using the algorithm proposed in [45].

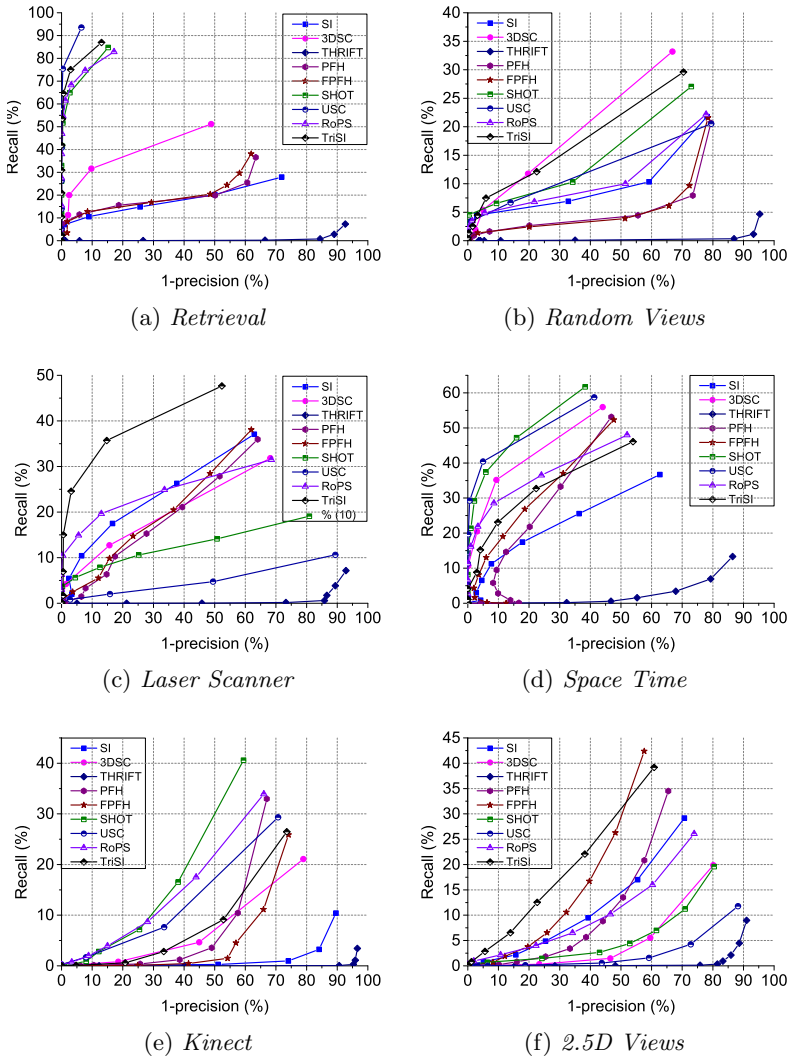
4 Performance Evaluation

4.1 Descriptiveness

We present the RPC results of these selected descriptors on the six datasets, as shown in Fig. 3.

***Retrieval* Dataset** The *Retrieval dataset* contains 18 scenes and 6 models. The scene meshes with three levels of noise are used in this experiment. USC achieves the best recall results, closely followed by TriSI, RoPS, and SHOT. 3DSC gives a moderate performance. Note that, the recall achieved by USC is much higher than 3DSC on the same dataset. This clearly demonstrates that the use of an LRF in USC not only reduces the memory requirements and the computational complexity of 3DSC, but also improves the matching accuracy of 3DSC [27]. Besides, FPFH, PFH, and SI have a similar performance, which is inferior to 3DSC.

***Random Views* Dataset** The *Random Views* dataset contains 36 scenes and 6 models. The scene meshes with three levels of noise are used in this experiment. 3DSC achieves the best performance, closely followed by TriSI. The next most performant descriptors are SHOT and USC. RoPS and SI achieve acceptable results, which are in fact much better compared to FPFH and PFH. As in the case of the *Retrieval* dataset, THRIFT gives the lowest scores. Note that, *Retrieval* and *Random Views* have the same models, with the major difference that *Random Views* contains occluded objects and clutter. Comparing the results in Figs. 3(a) and (b), three observations can be made. First, the recall on *Random Views* is significantly lower than the recall of *Retrieval* due to the more challenging conditions caused by occlusions and clutter. Second, when comparing the difference of the performance of each descriptor on these two datasets, USC, TriSI, RoPS, and SHOT have a larger drop compared to other descriptors. This is because these four descriptors are very sensitive to occlusions and clutter (see Section 4.2 under "Support Radius" and "Distance to the Image Boundary"). Third, the rankings of these descriptors on these two datasets are similar



(a) *Retrieval* (b) *Random Views*

(c) *Laser Scanner* (d) *Space Time*

(e) *Kinect* (f) *2.5D Views*

Fig. 3: Descriptiveness of the selected descriptors of Section 2 on the six datasets of Section 3.1 (Figure best seen in color).

except for SI, USC and 3DSC. USC is more suitable for the scenario of 3D shape retrieval, while 3DSC and SI are more suitable for 3D object recognition.

***Laser Scanner* Dataset** TriSI achieves the best results, showing a significant improvement compared to the other descriptors. RoPS and SI have a similar performance, followed by 3DSC, FPFH, and PFH. Note that, FPFH reduces the computational complexity of feature generation by an order of magnitude over PFH, while maintaining a similar performance in terms of feature matching

accuracy. SHOT produces moderate results, which are better than those of USC. 3DSC performs much better than USC on this dataset at the cost of an increased computational complexity and storage requirement.

***Space Time* Dataset** USC and SHOT outperform all the other descriptors, closely followed by 3DSC. It can be concluded that the shape context style descriptors (such as 3DSC and USC) are more suitable for applications with *Space Time*. The next performant descriptors are RoPS, TriSI, FPFH, and PFH. All these four descriptors produced a very close performance. SI achieves a moderate performance, with a lower recall compared to TriSI.

***Kinect* Dataset** SHOT achieves the best performance, followed by RoPS and USC. PFH and TriSI produced a similar performance. Note that, the recall achieved by FPFH is lower than PFH. Similarly, the recall obtained by 3DSC is lower compared USC. We also observe that SI and THRIFT do not work well on this dataset, mainly due to its high sensitivity to noise (as shown in Section 4.2). SI and THRIFT are therefore more suitable for applications on images with low noise and high resolution (e.g., *Laser Scanner*).

***2.5D Views* Dataset** TriSI gives the best results, followed by FPFH. PFH produced a much lower score compared to FPFH although the former requires more computational time. SI performs slightly better than PFH, which is followed by RoPS. 3DSC and SHOT have a very close performance, achieving a relatively low recall. Besides, the scores of USC and THRIFT are amongst the lowest. Note that, both *2.5D Views* and *Laser Scanner* were acquired with Minolta Vivid 910. The major difference between the two datasets is that *Laser Scanner* contains both occlusions and clutter while *2.5D Views* contains only occlusions. Compared to the results reported on *Laser Scanner* (Fig. 3(c)), several observations can be drawn. First, the rankings of these descriptors are similar on the two datasets. TriSI gives the best results, while SHOT, USC, and THRIFT achieve relatively low scores. Second, the superior performance of TriSI is more significant on *Laser Scanner* compared to *2.5D Views*. Third, FPFH performs better than PFH, and 3DSC achieved a better performance compared to USC on both of these two datasets. Fourth, RoPS is more suitable for object recognition compared to 3D modeling, FPFH and PFH are more suitable for 3D modeling compared to object recognition.

Descriptiveness Overall Performance In order to directly compare the performance of these descriptors on each dataset, we calculate the recall at the precision of 50% (denoted by $recall_{0.5p}$). $recall_{0.5p}$ is an established methodology previously adopted in [7] for the evaluation of a descriptor with a single number. The $recall_{0.5p}$ results of all these descriptors on the six datasets are presented in Table 2. We also present the average and median $recall_{0.5p}$ of the descriptors over all datasets. Several conclusions can be summarized as follows.

Table 2: The recall at 50% precision of the descriptors of Section 2 on the six datasets of Section 3.1. The best performance is reported in bold face, and the highest results for each dataset are shown in blue (Table best seen in color).

Dataset \ Descriptor	SI	3DSC	THRIFT	PFH	FPFH	SHOT	USC	RoPS	TriSI
Retrieval	20.2	51.2	0.1	19.9	21.5	84.8	**93.5**	82.9	87.0
Random Views	9.1	**25.6**	0.2	4.2	3.8	17.1	14.3	9.9	22.2
Laser Scanner	31.6	25.2	0.1	27.0	29.5	14.0	4.8	28.0	**46.9**
Space Time	31.3	56.0	1.0	53.1	52.3	**61.7**	58.7	47.2	44.4
Kinect	0.3	7.1	0	4.3	1.1	**30.0**	17.3	22.0	8.2
2.5D Views	14.5	2.6	0	13.0	29.3	3.9	1.0	11.7	**31.0**
Average	17.8	27.9	0.2	20.2	23.0	35.3	31.6	33.6	**40.0**
Median	17.4	25.4	0.1	16.4	25.4	23.6	15.8	25.0	**37.7**

First, TriSI, SHOT, and RoPS are amongst the best descriptors. Specifically, SHOT achieves the best performance on the *Space Time* and *Kinect* datasets. TriSI performs best on the *Laser Scanner* and *2.5D Views* datasets. Overall, TriSI has the highest average and median recall across all these datasets. It outperforms SHOT by a large margin, with average values of $recall_{0.5p}$ being 40.0 and 35.3, respectively. In contrast, THRIFT is the descriptor with the lowest performance on all these datasets.

Second, the performance of these descriptors depends on the dataset. It is clear that USC, 3DSC, TriSI, RoPS, and SHOT are the descriptors which produce the best performance on high resolution datasets (i.e., *Retrieval* and *Random Views*). Besides, SHOT, USC, and RoPS have a relatively better performance compared to all the others when tested on low resolution datasets (i.e., *Space Time* and *Kinect*). Moreover, TriSI, RoPS, SI, and FPFH are the top descriptors on the medium-level resolution datasets (i.e., *Laser Scanner* and *2.5D Views*).

Third, PFH, FPFH, TriSI, SI, and 3DSC generally show a more stable performance across datasets compared to all the others. In contrast, the performance of SHOT and USC varies significantly, as revealed by the large differences between their average and median values of $recall_{0.5p}$. This conclusion corroborates with the results in [14] and [23].

4.2 Robustness

In this section, we present the $recall_{0.5p}$ results of these descriptors with respect to different variations, as shown in Fig. 4. In this paper, we only present experimental results on the *Laser Scanner* dataset due to the limited number of pages. Note that, *Laser Scanner* is one of the most frequently used datasets in 3D computer vision [1, 44, 38, 9, 10].

Support Radius The support radius affects both the feature's descriptiveness and its robustness to occlusions and clutter [10]. Two major observations can

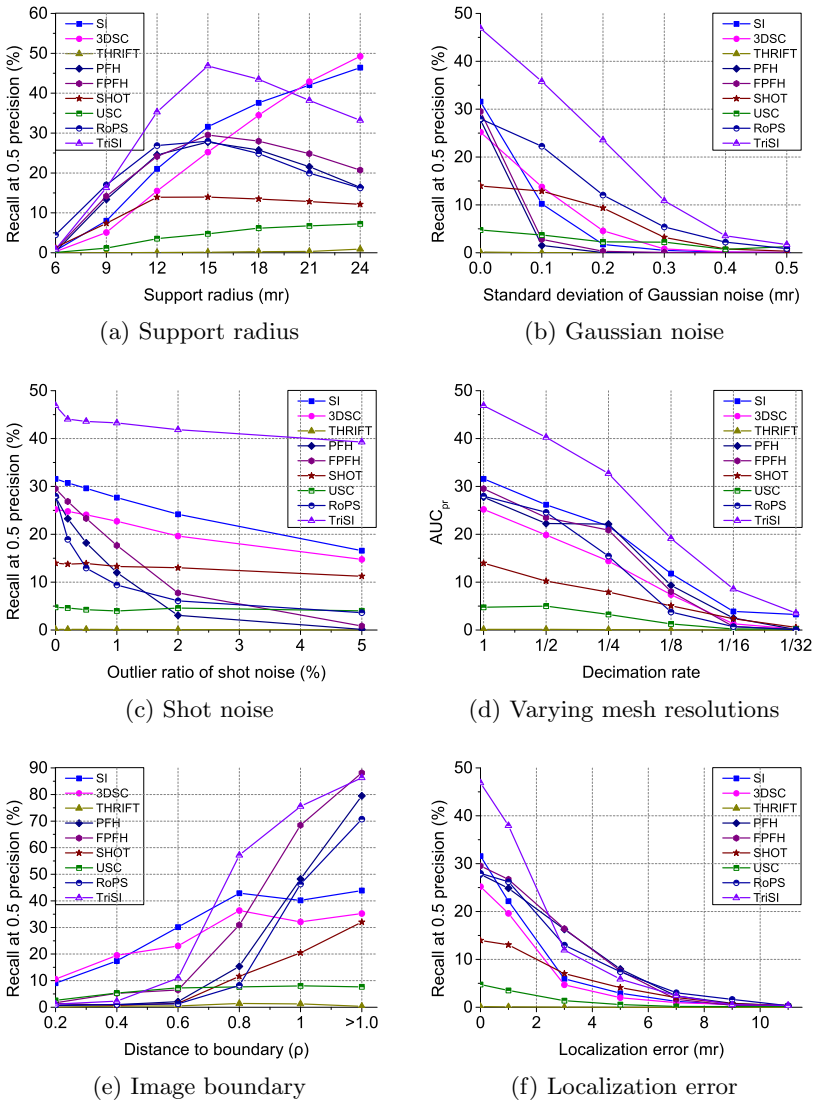
be made from the results in Fig. 4(a). First, the recall results of TriSI, FPFH, PFH, RoPS, and SHOT improve rapidly when the support radius is increased. Their performance reaches the peak value with a support radius of about 15mr. Their performance then decreases with an increase in the support radius. This is because these descriptors are highly sensitive to occlusions and clutter (as further demonstrated in Section 4.2 "Distance to the Image Boundary"). They produce the best performance when an optimal tradeoff is achieved between their descriptiveness and sensitivity. Second, for the descriptors which are less sensitive to occlusions and clutter (e.g., SI, 3DSC, and USC), their performance increases consistently with an increase in the support radius. This is because, the major factor which influences their performance is the encapsulated information of the underlying local surface rather than occlusions and clutter, as further explained in Section 4.2 "Distance to the Image Boundary".

Gaussian Noise The performance of all descriptors decreases very rapidly when the standard deviation of the Gaussian noise increases. USC is the most robust descriptor with respect to Gaussian noise, its value is very stable under different levels of noise. RoPS, TriSI, and SHOT also have acceptable robustness with respect to Gaussian noise. On the other hand, SI, THRIFT, PFH, and FPFH are very sensitive to Gaussian noise, their recall drops significantly when the standard deviation of Gaussian noise increases to 0.1mr. This is because they rely on first-order surface derivatives (i.e., surface normal), which are prone to noise.

Shot Noise TriSI is highly robust to shot noise, achieving a high recall (close to 40%) with an outlier ratio of shot noise of 5%. USC and SHOT are also very robust to shot noise, their performances drop slowly when the shot noise increases. The other descriptors are more affected by shot noise. RoPS achieves similar results compared to PFH with low levels of shot noise. It then outperforms PFH when the level of shot noise is high. Both FPFH and PFH are highly sensitive to shot noise, their performance deteriorates dramatically even with a low level of shot noise. From Figs. 4(b) and (c), it is clear that TriSI, USC, and SHOT are robust while PFH and FPFH are sensitive to both Gaussian and shot noise.

Varying Mesh Resolutions The recall of all these descriptors decreases as the level of mesh decimation increases. TriSI has the best performance under all levels of mesh decimation. PFH, FPFH, SI, TriSI are robust to varying mesh resolutions. Their drop in performance with respect to varying mesh resolutions is smaller compared to other descriptors. In contrast, THRIFT and USC are sensitive to varying mesh resolutions.

Distance to the Image Boundary The performance of TriSI is significantly boosted by eliminating points which are close to the image boundary. Specifically, the $recall_{0.5p}$ is increased from about 10% to about 60% by removing points with



(a) Support radius (b) Gaussian noise

(c) Shot noise (d) Varying mesh resolutions

(e) Image boundary (f) Localization error

Fig. 4: Robustness of the selected descriptors of Section 2 on the *Laser Scanner* dataset (Figure best seen in color).

distances less than 0.8ρ to the boundary. Similarly, the recall results of FPFH, PFH, RoPS, and SHOT are also significantly improved by removing boundary points. In contrast, SI, 3DSC, and USC are more robust to boundary points. SI and 3DSC achieve the best performance compared to all other descriptors when tested on keypoints with distances less than 0.7ρ to the boundary. Since the points close to the boundary include occlusions and clutter (as shown in Fig. 2) it can be concluded that TriSI, FPFH, PFH, RoPS, and SHOT are sensitive

to occlusions and clutter. In contrast, SI, 3DSC, and USC are very robust to occlusions and clutter. This is consistent with the conclusions drawn in Section 4.2 "Support Radius".

Keypoint Localization Error The recall decreases with increasing keypoint localization errors. The performance of TriSI, SI, and 3DSC drops faster than all the other descriptors, especially at keypoints with small localization errors. This indicates that these three descriptors are very sensitive to the accuracy of the keypoint localization. For keypoints with localization errors less than 3mr, the superior performance of TriSI is highly significant compared to the other descriptors. For keypoints with localization errors of more than 5mr, TriSI, RoPS, FPFH, PFH produce a very close performance. Their recall is the highest compared to the other descriptors. Besides, SI, SHOT, and 3DSC achieve the second best performance.

5 Conclusions

This paper presents a comprehensive evaluation of 3D local feature descriptors on a variety of datasets. It can serve as a "User Guide" for the selection of the most appropriate feature descriptor in the area of 3D computer vision. The descriptiveness of these descriptors was tested on six datasets in different application contexts. Generally, TriSI achieved the best overall results in terms of recall. USC, RoPS, and SHOT also produce good scores on some of these datasets. The robustness of these descriptors was also evaluated with respect to a number of nuisances. TriSI, SHOT, and USC are very robust to both Gaussian noise and shot noise, while PFH and FPFH are highly sensitive. SI and 3DSC are very robust to the distance to image boundary, while TriSI, RoPS, SHOT, PFH, and FPFH are all very sensitive. Moreover, the performance of TriSI, SI, and 3DSC dropped significantly when the localization error of the keypoints increased.

While these descriptors perform well on high resolution datasets (collected using costly scanners), their performance is rather weak with data from low-cost sensors (e.g., Kinect). Research should therefore be directed towards the design of suitable descriptors for low resolution and high-level noise data, or the design of higher resolution and low-cost RGBD cameras. In this paper, feature descriptors are extracted from the randomly selected ground-truth corresponding points between the scene and model. Therefore, the affect of keypoint detection algorithm on the feature matching performance of feature descriptors is not considered. In order to better resemble real applications, we will test these descriptors in combination with different 3D keypoint detectors in our future work.

Acknowledgement. This research was supported in part by the National Natural Science Foundation of China under Grant 61471371, and in part by the Australian Research Council under Grants DE120102960 and DP110102166.

References

1. Mian, A., Bennamoun, M., Owens, R.: Three-dimensional model-based object recognition and segmentation in cluttered scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (2006) 1584–1601
2. Guo, Y., Wan, J., Lu, M., Niu, W.: A parts-based method for articulated target recognition in laser radar data. Optik - International Journal for Light and Electron Optics **124** (2013) 2727–2733
3. Bronstein, A., Bronstein, M., Guibas, L., Ovsjanikov, M.: Shape google: Geometric words and expressions for invariant shape retrieval. ACM Transactions on Graphics **30** (2011) 1–20
4. Guo, Y., Sohel, F., Bennamoun, M., Wan, J., Lu, M.: An accurate and robust range image registration algorithm for 3D object modeling. IEEE Transactions on Multimedia **16** (2014) 1377–1390
5. Lei, Y., Bennamoun, M., Hayat, M., Guo, Y.: An efficient 3D face recognition approach using local geometrical signatures. Pattern Recognition **47** (2014) 509–524
6. Bennamoun, M., Guo, Y., Sohel, F.: 2D and 3D feature selection for face recognition. Encyclopedia of Electrical and Electronics Engineering, under review (2014)
7. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 1615–1630
8. Guo, Y., Bennamoun, M., Sohel, F., Lu, M., Wan, J.: 3D object recognition in cluttered scenes with local surface features: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, in press (2014)
9. Tombari, F., Salti, S., Di Stefano, L.: Performance evaluation of 3D keypoint detectors. International Journal of Computer Vision **102** (2013) 198–220
10. Guo, Y., Sohel, F., Bennamoun, M., Lu, M., Wan, J.: Rotational projection statistics for 3D local surface description and object recognition. International Journal of Computer Vision **105** (2013) 63–86
11. Bronstein, A., Bronstein, M., Bustos, B., et al.: SHREC 2010: Robust feature detection and description benchmark. In: Eurographics Workshop on 3D Object Retrieval. Volume 2. (2010) 6
12. Shah, S.A.A., Bennamoun, M., Boussaid, F., El-Sallam, A.: A novel local surface description for automatic 3D object recognition in low resolution cluttered scenes. In: IEEE International Conference on Computer Vision Workshops. (2013) 638–643
13. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. International Journal of Computer Vision **37** (2000) 151–172
14. Restrepo, M.I., Mundy, J.L.: An evaluation of local shape descriptors in probabilistic volumetric scenes. In: British Machine Vision Conference. (2012) 1–11
15. Guo, Y., Bennamoun, M., Sohel, F., Lu, M., Wan, J.: An integrated framework for 3D modeling, object detection and pose estimation from point-clouds. IEEE Transactions on Instrumentation and Measurement, in press (2014)
16. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.: A comparison of affine region detectors. International Journal of Computer Vision **65** (2005) 43–72
17. Moreels, P., Perona, P.: Evaluation of features detectors and descriptors based on 3D objects. In: 10th IEEE International Conference on Computer Vision. Volume 1. (2005) 800–807
18. Moreels, P., Perona, P.: Evaluation of features detectors and descriptors based on 3D objects. International Journal of Computer Vision **73** (2007) 263–284

19. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. In: IEEE Conference on Computer Vision and Pattern Recognition. Volume 2. (2003) II–257
20. Boyer, E., Bronstein, A., Bronstein, M., et al.: SHREC 2011: Robust feature detection and description benchmark. In: Eurographics Workshop on Shape Retrieval. (2011) 79–86
21. Salti, S., Tombari, F., Stefano, L.: A performance evaluation of 3D keypoint detectors. In: International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission. (2011) 236–243
22. Alexandre, L.A.: 3D descriptors for object and category recognition: a comparative evaluation. In: Workshop on Color-Depth Camera Fusion in Robotics at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). (2012)
23. Kim, H., Hilton, A.: Evaluation of 3D feature descriptors for multi-modal data registration. In: International Conference on 3D Vision. (2013) 119–126
24. Salti, S., Petrelli, A., Tombari, F., Di Stefano, L.: On the affinity between 3D detectors and descriptors. In: 2nd International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT). (2012) 424–431
25. Johnson, A.E., Hebert, M.: Using spin images for efficient object recognition in cluttered 3D scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence **21** (1999) 433–449
26. Frome, A., Huber, D., Kolluri, R., Bülow, T., Malik, J.: Recognizing objects in range data using regional point descriptors. In: 8th European Conference on Computer Vision. (2004) 224–237
27. Tombari, F., Salti, S., Di Stefano, L.: Unique shape context for 3D data description. In: ACM Workshop on 3D Object Retrieval. (2010) 57–62
28. Guo, Y., Bennamoun, M., Sohel, F., Wan, J., Lu, M.: 3D free form object recognition using rotational projection statistics. In: IEEE 14th Workshop on the Applications of Computer Vision. (2013) 1–8
29. Guo, Y., Sohel, F., Bennamoun, M., Wan, J., Lu, M.: A novel local surface feature for 3D object recognition under clutter and occlusion. Information Sciences, in press (2014)
30. Flint, A., Dick, A., Van den Hengel, A.: Local 3D structure recognition in range images. IET Computer Vision **2** (2008) 208–217
31. Rusu, R.B., Blodow, N., Marton, Z.C., Beetz, M.: Aligning point cloud views using persistent feature histograms. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. (2008) 3384–3391
32. Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (FPFH) for 3D registration. In: IEEE International Conference on Robotics and Automation. (2009) 3212–3217
33. Tombari, F., Salti, S., Di Stefano, L.: Unique signatures of histograms for local surface description. In: European Conference on Computer Vision. Springer (2010) 356–369
34. Salti, S., Tombari, F., Stefano, L.D.: SHOT: Unique signatures of histograms for surface and texture description. Computer Vision and Image Understanding, In press (2014)
35. Taati, B., Greenspan, M.: Local shape descriptor selection for object recognition in range data. Computer Vision and Image Understanding **115** (2011) 681–694
36. Lo, T., Siebert, J.: Local feature extraction and matching on range images: 2.5D SIFT. Computer Vision and Image Understanding **113** (2009) 1235–1250

37. Bayramoglu, N., Alatan, A.: Shape index SIFT: Range image recognition using local features. In: 20th International Conference on Pattern Recognition. (2010) 352–355
38. Bariya, P., Novatnack, J., Schwartz, G., Nishino, K.: 3D geometric scale variability in range images: Features and descriptors. International Journal of Computer Vision **99** (2012) 232–255
39. Pottmann, H., Wallner, J., Huang, Q.X., Yang, Y.L.: Integral invariants for robust geometry processing. Computer Aided Geometric Design **26** (2009) 37–60
40. Albarelli, A., Rodola, E., Torsello, A.: Loosely distinctive features for robust surface alignment. In: European Conference on Computer Vision. (2010) 519–532
41. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (1992) 239–256
42. Zaharescu, A., Boyer, E., Horaud, R.: Keypoints and local descriptors of scalar functions on 2D manifolds. International Journal of Computer Vision **100** (2012) 78–98
43. Rusu, R.B., Cousins, S.: 3D is here: Point cloud library (PCL). In: IEEE International Conference on Robotics and Automation. (2011) 1–4
44. Mian, A., Bennamoun, M., Owens, R.: On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes. International Journal of Computer Vision **89** (2010) 348–361
45. Chen, X., Schmitt, F.: Intrinsic surface properties from surface triangulation. In: European Conference on Computer Vision. (1992) 739–743