



# Multi-attribute statistics histograms for accurate and robust pairwise registration of range images



Jiaqi Yang, Qian Zhang, Zhiguo Cao\*

Guangdong HUST Industrial Technology Research Institute, Guangdong Province Key Lab of Digital Manufacturing Equipment, Image Processing and Intelligent Control Key Laboratory of Education Ministry of China, School of Automation, Huazhong University of Science and Technology, Wuhan, China

## ARTICLE INFO

### Article history:

Received 8 August 2016

Revised 20 January 2017

Accepted 5 April 2017

Available online 14 April 2017

Communicated by Bin Fan

### Keywords:

Local reference axis

Multi-attribute

Transformation estimation

Feature matching

Range image registration

## ABSTRACT

Registration of range images based on local shape features is widely adopted due to its validated effectiveness and robustness, while most existing local shape descriptors struggle to simultaneously achieve a pleasurable and balanced performance in terms of distinctiveness, robustness and time efficiency. This paper proposes a novel representation of 3D local surfaces, called multi-attribute statistics histograms (MaSH), for automatic registration of range images. MaSH comprises both spatial and geometric information characterizations. The characterization of spatial information is achieved via radial partitions in the 3D local support volume around the keypoint based on a local reference axis (LRA), creating a set of subspaces. While the encoding the shape geometry is performed by calculating statistical histograms of multiple faint correlated geometric attributes (i.e., local depth, normal deviation, and surface variation angle) for each subspace, so as to achieve information complementarity. Then, a robust rigid transformation estimation algorithm named 2-point based sample consensus with global constrain (2SAC-GC) is presented to tackle the problem of calculating an optimal transformation from the correspondence set with outliers. Finally, a coarse-to-fine registration method based on MaSH and 2SAC-GC is proposed for aligning range images. Experiments on both high-resolution (Laser Scanner) and low-resolution (Kinect) datasets report that, our method achieves a registration accuracy of 90.36% and 80.39% on the two datasets, respectively. It also exhibits strong robustness against noise and varying mesh resolutions. Furthermore, feature matching experiments show the over-all superiority of the proposed MaSH descriptor against the state-of-the-arts including the spin image, snapshots, THRIFT, FPFH, RoPS, LFSH and RCS descriptors.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

There has been a long research history on the computer vision tasks based on 2D images. Despite their notable progress, there are still numerous obstacles left due to the essential shortcomings of 2D images, e.g., sensitivities to scale, rotation and illumination. In contrast, range images, as the 3D representation of the real world, show many favorable traits against 2D images, such as providing more geometrical (depth) information and scale invariance. Accordingly, vision tasks based on range images have attracted growing interests. The development of numerous low-cost 3D sensors, e.g., Microsoft Kinect and Intel RealSense, has further promoted such trend in recent years. Pairwise registration of range images is a fundamental and crucial step in many 3D vision tasks,

such as 3D object recognition [1], surface registration [2], 3D modeling [3], cultural heritage [4] and robotics [5].

Given two range images, namely a source one and a target one, the aim of pairwise range image registration is to transform the coordinate of the source to that of the target. The challenges are many fold, e.g., unknown initial poses, noise injected by the acquisition system, varying mesh resolutions (point densities) caused by different scanning distances, holes and small overlaps. Some applications, such as robotics, even have strict demands on time efficiency. To solve above difficulties, many efforts have been made in the literature. A matured pipeline for range image registration is a coarse-to-fine strategy [3,6–8], where the coarse registration provides an initial guess for the fine registration as well as accelerates the registration process. The most typical fine registration methods are arguably the iterative closest point (ICP) algorithm and its variants [9–11]. Since the existing fine registration methods are already able to obtain fairly accurate results [12], the real challenging turns out to be the coarse registration problem, which has attracted many interests lately. Feature-based

\* Corresponding author.

E-mail addresses: [jqyang@hust.edu.cn](mailto:jqyang@hust.edu.cn) (J. Yang), [hangfanzq@163.com](mailto:hangfanzq@163.com) (Q. Zhang), [zgcao@hust.edu.cn](mailto:zgcao@hust.edu.cn) (Z. Cao).

coarse registration methods are mostly used nowadays owing to their validated effectiveness and robustness [3,6,13–17], their employed features can be classified into local and global features. Owing to the superiority of local features against global features in terms of distinctiveness and robustness to clutter, occlusion, and missing regions [12], local features are more suitable for generating correspondences between range images.

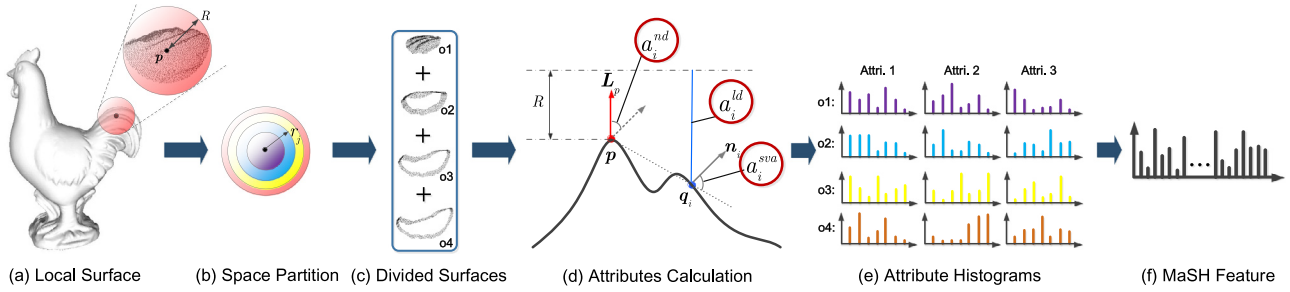
For local feature-based methods, local 3D descriptor and transformation estimation play two critical roles because the former one directly affects the quality of the established correspondences between two range images, and the later one is amenable to calculate a reasonable transformation from a correspondence set with outliers. In the following, we briefly review the existing local feature descriptors and 3D transformation estimation methods in the literature. Johnson and Hebert [6,18] proposed a spin image descriptor for surface matching and object recognition. They represented the neighboring points of a keypoint with a cylindrical coordinate frame, and encoded the point distribution with a 2D image, named a spin image. Spin image is one of the most cited descriptors, but still suffers from limited descriptiveness and poor robustness. Several variants of spin image are later proposed, including the spherical spin image [19] and Tri-Spin-Image (TriSI) descriptor [20]. Frome et al. [21] proposed a 3D shape context (3DSC) descriptor, which is an extension of the 2D shape context method [22]. They first divided the local spherical volume around the keypoint into a set of subspaces based on a local reference frame (LRF), and then calculated the weighted number of points in each subspace. Flint et al. [15,23] first calculated the normal deviations between the normals of the neighbors and that of the keypoint, and then represented the normal deviation distribution with a 1D statistical histogram, called THRIFT. Similarly, Rusu et al. [7,8] proposed the point feature histograms (PFH) and fast point feature histograms (FPFH) for characterizing normal attributes. Malassiotis and Srinivas [24] first built an LRF centered at the keypoint by performing covariance analysis on the neighboring points, they then captured a local depth image at a virtual view point in the z-axis of the LRF, generating a “snapshots” descriptor. Guo et al. [25] proposed a rotational projection statistics (RoPS) descriptor, which is the concatenation of the statistics calculated for each 2D projection map after rotating the local surface multiple times. Following the multi-view mechanism in RoPS, a rotation contour signature (RCS) [26] based on 2D contour feature representation is proposed in [26], delivering better robustness than RoPS. Shah et al. [27,28] proposed a 3D vorticity (3D-Vor) descriptor based on the vorticity of the normal vector field computed at each neighboring point. 3D-Vor particularly addresses low-resolution range image registration scenario. Recently, Yang et al. [29] proposed a local feature statistics histograms (LFSH) descriptor to achieve fast point cloud registration using several low-dimensional and efficient geometrical features. Although there has been a variety of descriptors at present, most of them still fail to achieve a balanced and pleasant performance. For instance, spin image and LFSH exhibit limited descriptiveness [30]; PFH and FPFH are sensitive to noise [3]; snapshots is prone to rotation [25]; RoPS demands mesh representation and is computationally expensive [31]. However, these are all major concerns for a local feature-based registration algorithm in practice.

In terms of transformation estimation, the random sample consensus (RANSAC) [32] is a commonly-used method in both 2D and 3D registration cases, and shows some specific advantages, e.g., low implementation complexity and very few tunable parameters, against other approaches such as 3D Hough transform [33], the rigidity constrain framework [34] and voting-based approach [35]. In 3D registration case, owing to that additional geometry information and/or constraints can be explored, many variants of

RANSAC are thereupon proposed to either improve its time efficiency and/or robustness. Rusu et al. [8] proposed a sample consensus initial alignment (SAC-IA) algorithm by selecting a set of sample correspondences with distance constrain from the initial correspondence set, and iteratively computing the optimal transformation that yields to best error metric (i.e., the Huber penalty measure). The SAC-IA algorithm reduces some unreasonable iterations by judging the qualification of current sampled correspondences according to a distance constrain, while it still struggles to take effect in cases with severe outliers. Guo et al. [36] recently proposed a 1-point RANSAC (1P-RANSAC) algorithm to reduce the high computational complexity of RANSAC. To be specific, they first calculated an LRF for each keypoint in both range images, and then iteratively sampled a single correspondence from the correspondence set for the calculation of the optimal transformation. Owing to that two corresponding LRFs are suffice to compute a rigid transformation [37], the 1P-RANSAC algorithm only needs one correspondence at each iteration in opposite to three in the classical RANSAC algorithm. However, it is sensitive to LRF calculation errors and high percentage of outliers in the initial correspondence set.

In these regards, we propose a novel local shape descriptor called multi-attribute statistics histograms (MaSH) together with a new transformation estimation algorithm named 2-point based sample consensus with global constrain (2SAC-GC), for accurate and robust pairwise registration of range images. MaSH contains both spatial and geometric information descriptions. The spatial information is encoded via partitions along the radial direction in the local 3D support volume based on an LRA. Employing 1D spatial information is a compromise among descriptiveness, robustness and efficiency, since more dimensional spatial information (e.g., the azimuth dimension) requires an LRF rather than an LRA. However, establishing a repeatable LRF in the local surface is far more challenging (i.e., needs to calculate three unique axes rather than one) and needs more time consumptions than an LRA [38,39]. As for geometric information description, statistics histogram of point attribute is a simple and effective feature representation as in [8,15,30]. Specifically, we calculate multiple (three in this paper) faint correlated point attributes for each neighboring point, and generate three statistic histograms for each subspace (partition). The idea behind multi-attribute description is to give a comprehensive geometric information encoding for the local surface, in opposite to traditional single attribute-based methods [15,18,25]. By concatenating these histograms in each subspace into a vector, a MaSH is generated. Then, we present a transformation estimation algorithm named 2-point based sample consensus with global constrain (2SAC-GC), which iteratively calculates the optimal transformation from two correspondences with two geometry constraints and a new discriminant criterion. Finally, a coarse-to-fine method is proposed for automatic pairwise registration of range images.

With the same purpose of local surface description as in our previous works [26,29], though, the designing techniques in the proposed MaSH descriptor are different and more advanced. First, compared with [29], MaSH integrates spatial information and histogram-weights to attain stronger discriminative power and resilience to self-occlusion. Second, in opposite to [26], MaSH is a histogram-based descriptor while [26] is signature-based. Further, MaSH does not rely on LRF for spatial information characterization. Third, the crafting scheme in MaSH particularly addresses the range image registration scenario, i.e., self-occlusion, missing region and holes, whereas [26,29] mainly consider complete local shapes. Comparative results (Section 4.3) clearly demonstrate that the proposed MaSH descriptor outperforms our previous works [26,29] on all the datasets in both feature matching and range image registration performance.



**Fig. 1.** The MaSH feature extraction pipeline. For clarity, we set 4 partitions along the radial direction for subspace generation, see (b) and (c). The attributes calculation, see (d), is illustrated in 2D, where the attributes are highlighted with red circles. Please refer to the text for the representations of variables. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 1.1. Contributions

The contributions of this paper are summarized as follows:

- We present a MaSH descriptor that simultaneously achieves high descriptiveness, robustness and time efficiency for the purpose of local surface representation and shape matching.
- We propose a robust rigid transformation estimation algorithm, i.e., the 2SAC-GC, for the 3D correspondence set contaminated by outliers.
- A coarse-to-fine pairwise registration method for range images is proposed, which is accurate and robust to many nuisances including noise and varying mesh resolutions.

### 1.2. Organization

The reminder of this paper is structured as follows. Section 2 gives the details of the MaSH descriptor. Section 3 presents the proposed coarse-to-fine registration method for range images. Experimental results including a detailed qualitative and quantitative analysis and comparisons with the state-of-the-arts are presented in Section 4. Finally, conclusions and future work are drew in Section 5.

## 2. Local surface representation with MaSH

The proposed methodology for computing a MaSH descriptor from a sphere-cropped local shape in a range image is depicted in Fig. 1. Briefly stated, a sphere is centered at a given keypoint to crop the local surface. Space partition along the radial direction is thereupon operated to separate the local surface into several fractions. The motivation behind is to exploit radial spatial information. Then, for each radius neighbor, three attributes including normal deviation, local depth and surface variation angle based on an LRA are respectively computed for feature description. These attributes would be further represented using statistical histograms with different weights. The final MaSH descriptor is the integration of these weighted-histograms. The details of each step are given below.

### 2.1. LRA construction and space partition

Given a range image  $\mathcal{R}$  and a keypoint  $\mathbf{p}$ , the radius neighbors of  $\mathbf{p}$  is calculated as  $\mathcal{Q} = \{\mathbf{q}_i : \|\mathbf{q}_i - \mathbf{p}\| \leq R\}$ , where  $R$  is the support radius of the 3D local spherical volume around  $\mathbf{p}$ . In order to encode the local spatial information, an LRA or LRF is required at the keypoint [25]. Although an LRF provides more spatial information (i.e., the azimuth spatial information), it is difficult to guarantee the LRF's robustness in cases of incomplete borders, holes, outliers and irregularly scanned points. In contrast, building a repeatable LRA is much easier and more efficient [38]. Here, we

employ the z-axis of the LRF proposals in [38,39] as our LRA. It is actually the normal vector of the keypoint calculated from a subset of the neighboring points. The benefit behind is reducing the impact brought by self-occlusion and incomplete borders. Note that although the employed LRA derives from an LRF, we do not compute LRF for our MaSH descriptor. To make the paper self-contained, we introduce the details of the LRA calculation. First, a small subset of  $\mathcal{Q}$  is extracted as  $\mathcal{Q}_L = \{\mathbf{q}_i^L : \|\mathbf{q}_i^L - \mathbf{p}\| \leq a \cdot R\}$ , where  $a$  is set to 0.33 as the authors suggested in [38,39]. Then, principle component analysis (PCA) is performed on  $\mathcal{Q}_L$  as:

$$\mathbf{C}(\mathcal{Q}_L) = \begin{bmatrix} \mathbf{q}_1^L - \bar{\mathbf{q}}^L \\ \vdots \\ \mathbf{q}_s^L - \bar{\mathbf{q}}^L \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{q}_1^L - \bar{\mathbf{q}}^L \\ \vdots \\ \mathbf{q}_s^L - \bar{\mathbf{q}}^L \end{bmatrix}, \quad (1)$$

where  $s$  is the number of points in  $\mathcal{Q}_L$ , and  $\bar{\mathbf{q}}^L$  represents the centroid of  $\mathcal{Q}_L$ . The eigenvector  $\mathbf{v}_{min}$  corresponding to the minimum eigenvalue of matrix  $\mathbf{C}(\mathcal{Q}_L)$  is then computed. We also calculate a vector  $-\mathbf{v}_{min}$  that is in opposite to  $\mathbf{v}_{min}$ , because the sign of  $\mathbf{v}_{min}$  is ambiguous [40]. To select a persistent vector from  $\mathbf{v}_{min}$  and  $-\mathbf{v}_{min}$ , we disambiguate the sign and calculate an LRA at  $\mathbf{p}$  as:

$$\mathbf{L}_p = \begin{cases} \mathbf{v}_{min}, & \text{if } \mathbf{v}_{min} \cdot \sum_{i=1}^s (\mathbf{p} - \mathbf{q}_i^L) \geq 0 \\ -\mathbf{v}_{min}, & \text{otherwise} \end{cases}. \quad (2)$$

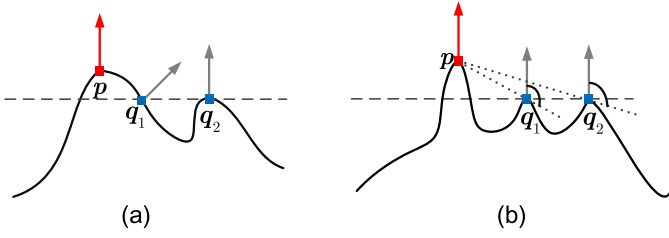
With an LRA at  $\mathbf{p}$ , we then divide the local 3D volume into several subspaces  $\{o_1, o_2, \dots, o_{N_r}\}$  via partitions along the radial direction, as shown in Fig. 1(b). We denote the  $N_r$  radial divisions by  $r = \{r_1, r_2, \dots, r_{N_r}\}$ . In order to achieve a uniform space partition in a spherical volume, the  $j$ th radial division is calculated as:

$$r_j = R \sqrt[3]{\frac{j}{N_r}}. \quad (3)$$

It is worth noting that nonuniform partition would result to uneven distribution of points in the subspaces, leading the subspaces with less points to be more sensitive to the variance of mesh resolutions as well as other nuisances.

### 2.2. Geometric attributes calculation

Shape geometry is usually represented by various geometric attributes, including low-order attributes such as depth and boundary contours, and high-order attributes such like shape index and mean curvature [41]. Both categories of geometric attributes have manifested their validity in local shape description [18,24,42]. Concerning that a single geometric attribute only reflects the shape geometry from a certain aspect, and may meet with limited descriptiveness for complex shapes, we therefore employ multiple attributes for feature representation in this paper. Specifically, three attributes, i.e., local depth, normal deviation and surface variation angle, are used in MaSH as illustrated in Fig. 1(d). For a



**Fig. 2.** An illustration of information complementary inherent to the three attributes in MaSH. The two neighbors  $q_1$  and  $q_2$  of the keypoint  $p$  in (a) and (b) are with either different geometric and/or spatial information. In (a), the two neighbors with same local depth values are distinguished by their normal deviation values. In (b), they share same local depth and normal deviation values, but are distinguished by the surface variation angle attribute.

neighbor  $q_i$  of  $p$ , the three attributes at  $q_i$  are calculated as:

$$\begin{cases} a_i^{ld} = R + L_p \cdot (\mathbf{p} - \mathbf{q}_i) \\ a_i^{nd} = \arccos(\mathbf{n}_i \cdot \mathbf{L}_p) \\ a_i^{sva} = \arccos(\mathbf{n}_i \cdot (\mathbf{q}_i - \mathbf{p})) \end{cases}, \quad (4)$$

where  $\mathbf{n}_i$  denotes the normal vector of  $q_i$ , and  $a_i^{ld}$ ,  $a_i^{nd}$  and  $a_i^{sva}$  respectively represent the local depth, normal deviation and surface variation angle values at  $q_i$ . Note that the surface normal calculation method used in this paper is PCA-based [43], requiring only surface vertices for computation. The ranges of the three attributes are  $[0, 2R]$ ,  $[0, 180]$  and  $[0, 180]$ , respectively. Thus, each neighboring point is attached with three attributes that will be encoded to histograms later. The employed three attributes in MaSH have many merits. For instance, the unique geometric and spatial information of a neighboring point in the local surface can be well encoded as shown in Fig. 2. It would provide information complementary to a feature descriptor. Another advantage is their high computational efficiency as demonstrated in [15,24]. These superior traits of the three attributes definitely help us achieve a comprehensive and efficient description for the local surface (as demonstrated in Section 4).

### 2.3. MaSH generation

In order to obtain the attribute distribution information in each subspace, we then calculate a statistics histogram for each category of attribute in a subspace. That is, three histograms, i.e.,  $\mathbf{h}^{ld}$ ,  $\mathbf{h}^{nd}$ , and  $\mathbf{h}^{sva}$ , which involve local depth, normal deviation, and surface variation angle statistics information, are computed for a subspace  $o_j$ . Next, the sub-feature  $\mathbf{f}_j$  of MaSH, as the integration of the three statistics histograms in  $o_j$ , is defined as:

$$\mathbf{f}_j = w_j \{\mathbf{h}^{ld}, \mathbf{h}^{nd}, \mathbf{h}^{sva}\}, \quad (5)$$

with

$$w_j = \exp\left(\frac{1-j}{N_r}\right). \quad (6)$$

Remarkably,  $w_j$  is a weight for sub-feature  $\mathbf{f}_j$  that is related to the distance of subspace  $o_j$  to the keypoint  $p$ . As highlighted in [25,30], the points in the local surface that are near to the keypoint appear to be more robust to clutter and occlusion. We therefore let the subspace that is close to the keypoint contribute more to shape description. Finally, the MaSH descriptor  $\mathbf{f}$  is calculated as the concatenation of the sub-features, that is:

$$\mathbf{f} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{N_r}\}. \quad (7)$$

Hence, MaSH is composed of  $N_r \times (N_{ld} + N_{nd} + N_{sva})$  bins in total, where  $N_{ld}$ ,  $N_{nd}$  and  $N_{sva}$  denote the number of bins in  $\mathbf{h}^{ld}$ ,  $\mathbf{h}^{nd}$  and  $\mathbf{h}^{sva}$ , respectively.

**Table 1**

Parameter settings for the analysis of MaSH parameters.

	$N_{ld}$	$N_{nd}$	$N_{sva}$	$N_r$	$R$ (mr)
Fig. 3(a)	2–16	1	1	5	20
Fig. 3(b)	1	2–16	1	5	20
Fig. 3(c)	1	1	2–16	5	20
Fig. 3(d)	10	8	10	2–9	20
Fig. 3(e)	10	8	10	5	5–30

The key traits of the proposed MaSH descriptor are summarized as follows:

- The theoretical computational complexity of MaSH for a given local surface with  $k$  points is  $O(k)$ .
- MaSH explores the spatial information based on an LRA rather than an LRF, in opposite to many descriptors with spatial information such as snapshots and RoPS. LRA provides less dimensions of spatial information and may lead to a feature descriptor with limited descriptiveness, though, it shows strong robustness to common nuisances [38]. To compensate for descriptiveness, multiple faint attributes are employed for description to achieve a comprehensive representation.
- All the employed attributes are encoded with a repeatable LRA, and are endowed with invariance to rigid transformation and robustness to noise, clutter and occlusion inherent to the LRA. In addition, all the statistics histograms in MaSH are normalized to achieve robustness to varying mesh resolutions.

### 2.4. Analysis of MaSH parameters

There are five key parameters in MaSH, i.e., the number of bins in  $\mathbf{h}^{ld}$ ,  $\mathbf{h}^{nd}$  and  $\mathbf{h}^{sva}$ , the number of subspaces  $N_r$ , and the size of support radius  $R$ . In order to determine a reasonable parameter setting for MaSH, we test the feature matching performance of MaSH under different parameter settings using a popular criterion, i.e., the recall versus 1-precision curve (RPC) [25,30,31]. The dataset in this experiment is composed of 12 range images with 6 source ones and 6 target ones. The source range images are the Armadillo, Happy Buddha, Bunny, Asia Dragon, Dragon and Thai Statue models taken from the Stanford 3D scanning repository [44], whereas the target range images are their rotated copies with 0.3 mr Gaussian noise and  $\frac{1}{4}$  mesh decimation. Here, the unit mr represents the average length of the edges of all the triangles in the source range images. By changing the parameter to be analyzed and keeping other parameters constant as reported in Table 1, the RPC results are shown in Fig. 3(a)–(e).

One can see that, proper values of  $N_{ld}$ ,  $N_{nd}$ ,  $N_{sva}$  are respectively in the range of [10, 16], [8, 16] and [8, 16], since they achieve similar RPC performance in these scopes as shown in Fig. 3(a)–(c). Fig. 3(d) shows that the RPC performance improves as  $N_r$  increases from 2 to 5, and remains almost unchanged as  $N_r$  gets larger than 5. For the parameter  $R$ , we can see in Fig. 3(e) that the RPC performance achieves a dramatic promotion when  $R$  increases from 5 mr to 20 mr, and then improves in a slight manner as  $R$  further increases. Further, to validate the observations in Fig. 3 as well as examine the MaSH's robustness in a more challenging case, we present the results on the turning dataset with 0.5 mr Gaussian noise and  $\frac{1}{8}$  mesh decimation in Fig. 4. As expected, higher level of noise and mesh decimation leads to a clear drop of performance on MaSH under all parameter settings. Still, we find that the optimum parameter selection conclusion coincides with the result in Fig. 3. Based on above observations together with a trade-off among descriptor's descriptiveness, compactness and computational efficiency, we set  $N_{ld} = 10$ ,  $N_{nd} = 8$ ,  $N_{sva} = 10$ ,  $N_r = 5$  and  $R = 20$  mr in this paper.



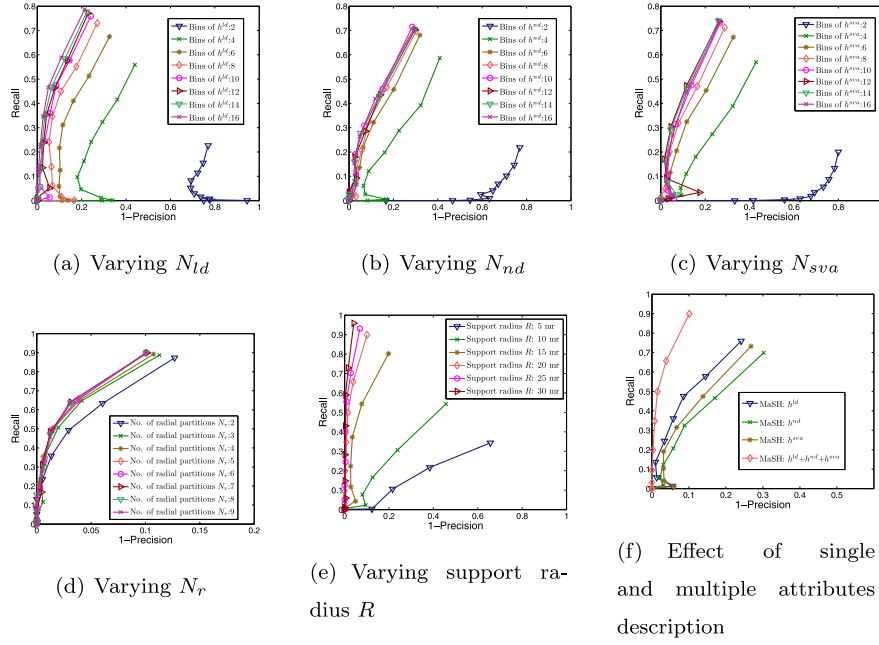


Fig. 3. (a)–(e) Analysis of MaSH parameters. (f) The effect of single and multiple attributes description on MaSH.

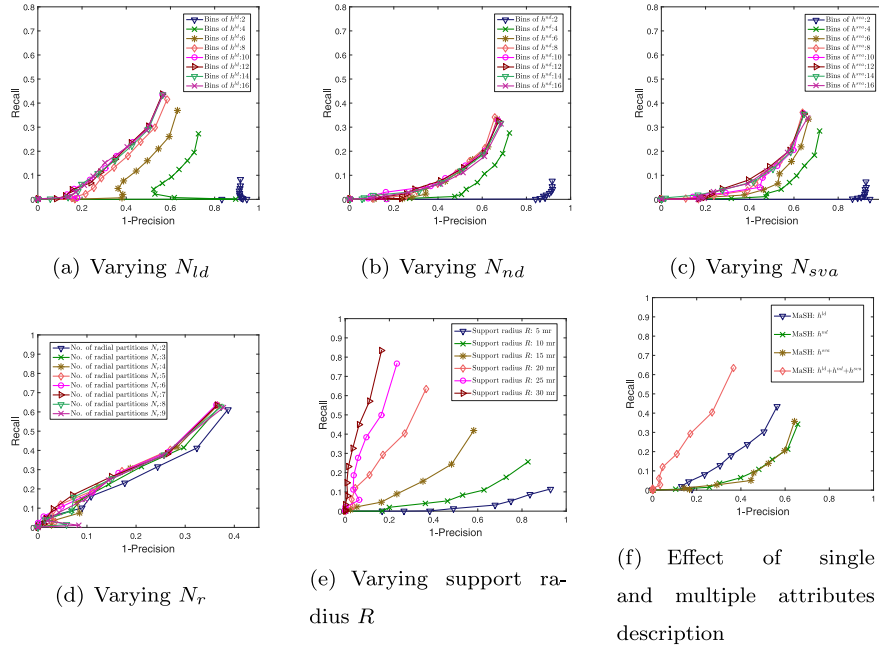


Fig. 4. Consistency validation of MaSH parameters (see Fig. 3) on a more challenging turning dataset with 0.5 mr Gaussian noise and  $\frac{1}{8}$  mesh decimation.

In addition, we also present the RPCs with and without multi-attribute description in MaSH as shown in Fig. 3(f) and Fig. 4(f). It is conspicuous that the integration of several faint correlated attributes makes a descriptor far more discriminative than the ones with a single attribute for shape description.

### 3. Coarse-to-fine pairwise range image registration

On the basis of the proposed MaSH descriptor, a coarse-to-fine method for pairwise range image registration is described in this section. It consists of four major modules: MaSH feature extraction, feature matching, robust transformation estimation via 2SAC-GC, and fine registration. Fig. 6 shows the MaSH-based range image registration pipeline.

#### 3.1. MaSH feature extraction

Let  $\mathcal{R}_s$  and  $\mathcal{R}_t$  respectively represent the input source and target range images. Before feature extraction, it is necessary to select a small subset of points (i.e., keypoints) in both range images to reduce computational costs and memory storage, because the raw acquired data by 3D scanners is usually of great redundancy. There is a vast number of keypoint detection methods for range images [37,45–47], which are either based on surface variations or random sampling. Although the keypoints detected by surface variation-based methods turn out to be more repeatable [46], most of them require expensive time costs. Since a fine registration step is prepared for further improving registration accuracy, we adopt a uniform sampling strategy for fast keypoint detection in this paper.

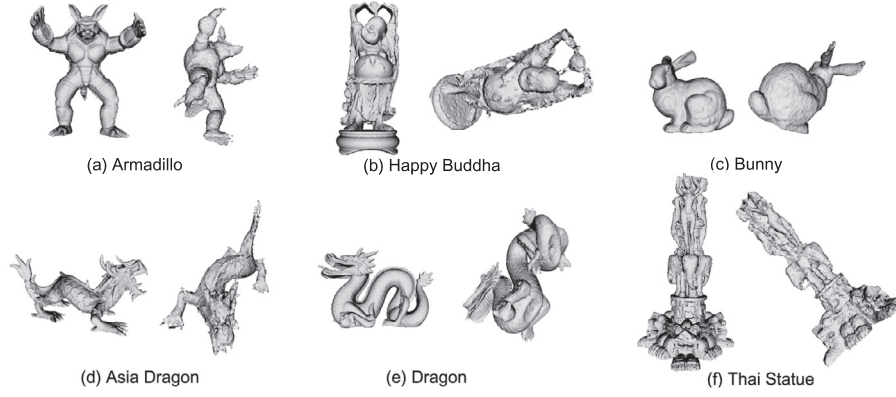


Fig. 5. The turning dataset. The left and right range images in each sub-figure are the source and target range images, respectively.

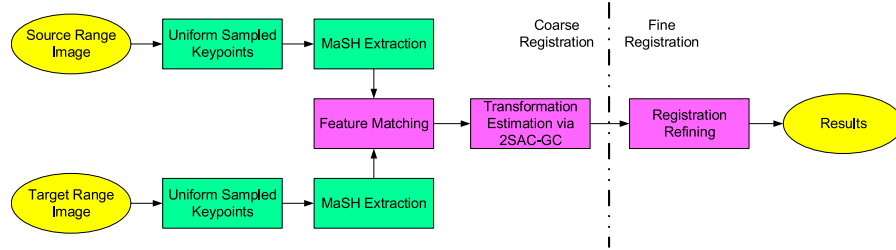


Fig. 6. Coarse-to-fine pairwise registration pipeline of range images based on MaSH.

First, both input range images are uniformly down-sampled using a real-time compression algorithm [48], creating another two low-resolution range images  $\tilde{\mathcal{R}}_s$  and  $\tilde{\mathcal{R}}_t$ . Then, the points in  $\mathcal{R}_s$  and  $\mathcal{R}_t$  that are closest to the points in  $\tilde{\mathcal{R}}_s$  and  $\tilde{\mathcal{R}}_t$  are defined as the keypoints, which are denoted by  $\{\mathbf{p}_1^s, \mathbf{p}_2^s, \dots, \mathbf{p}_{k_s}^s\}$  and  $\{\mathbf{p}_1^t, \mathbf{p}_2^t, \dots, \mathbf{p}_{k_t}^t\}$ , respectively.  $k$ -d tree is used here for efficient searching of closest points. At last, MaSH features for these keypoints are calculated as  $\{\mathbf{f}_1^s, \mathbf{f}_2^s, \dots, \mathbf{f}_{k_s}^s\}$  and  $\{\mathbf{f}_1^t, \mathbf{f}_2^t, \dots, \mathbf{f}_{k_t}^t\}$ .

### 3.2. Feature matching

Having calculated a set of features for  $\mathcal{R}_s$  and  $\mathcal{R}_t$ , we are left with the task of generating correspondences between them. Thus, a source feature  $\mathbf{f}_i^s$  finds its closest feature  $\mathbf{f}_i^t$  in  $\{\mathbf{f}_1^t, \mathbf{f}_2^t, \dots, \mathbf{f}_{k_t}^t\}$  as:

$$\mathbf{f}_i^t = \arg \min_{n=1,2,\dots,k_t} (\|\mathbf{f}_i^s - \mathbf{f}_n^t\|). \quad (8)$$

The associated point pair  $(\mathbf{p}_i^s, \mathbf{p}_i^t)$  of  $(\mathbf{f}_i^s, \mathbf{f}_i^t)$  is regarded as a point-to-point correspondence.  $k$ -d tree is also used here to accelerate feature matching process. Note that  $L_2$  norm is employed for measuring the similarity of two feature vectors, as in [8,30]. Therefore,  $k_s$  correspondences can be generated. However, only a portion of them is supposed to be correct because the input range images are usually partially overlapped. We then rank the  $k_s$  correspondences based on the similarity measured by  $L_2$  norm between the associated features of a correspondence, and keep the former  $k$  ones as the initial correspondence set  $\mathcal{C}$ .

### 3.3. Robust transformation estimation via 2SAC-GC

Thus far, a correspondence set  $\mathcal{C}$  is established between  $\mathcal{R}_s$  and  $\mathcal{R}_t$ . The next task is to estimate a rigid transformation from  $\mathcal{C}$  so as to transform  $\mathcal{R}_s$  to  $\mathcal{R}_t$ . Popular techniques include the RANSAC algorithm [32] and its variants [8,36]. Unfortunately, they still suffer from weak robustness to high ratio of outliers and/or high time consumption [37]. Here, we propose a 2-point based sample

consensus with global constrain (2SAC-GC) algorithm for robust transformation estimation.

The 2SAC-GC algorithm iteratively estimates the optimal transformation from  $\mathcal{C}$  in pre-defined  $N_{iter}$  iterations. At the  $i$ th iteration, two correspondences  $c_1^i$  and  $c_2^i$  are randomly selected from  $\mathcal{C}$ . Theoretically, three points are required for calculating a transformation in 3D case. Nevertheless, if the points are orientated, i.e., attached with LRAs in our method, two points are sufficient [49]. Note that, the LRAs of each keypoint have been calculated in the feature extraction process, and no other LRA calculations are needed here. The benefit of using two correspondences is reducing the computational complexity  $O(k^3)$  to  $O(k^2)$ , in opposite to classical RANSAC. Before calculating a transformation using  $c_1^i$ ,  $c_2^i$  and their corresponding LRAs, we enforce two additional constraints based on Euclidean distance and angle consistencies on the two correspondences as:

$$\|\mathbf{p}_1^s - \mathbf{p}_2^s\| - \|\mathbf{p}_1^t - \mathbf{p}_2^t\| \leq \sigma_d \quad (9)$$

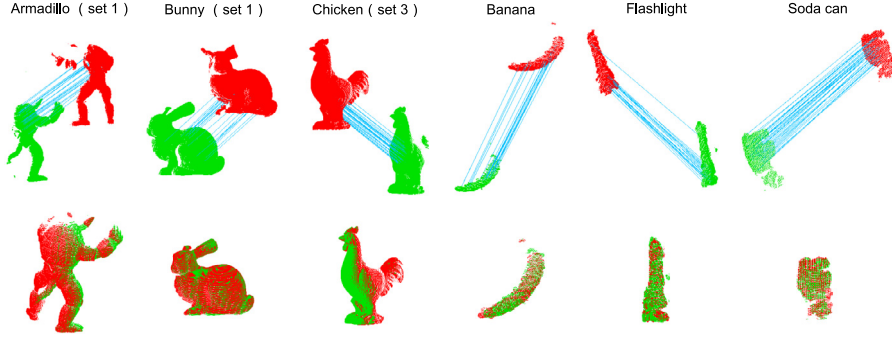
and

$$|\arccos(\mathbf{L}_1^s \cdot \mathbf{L}_2^s) - \arccos(\mathbf{L}_1^t \cdot \mathbf{L}_2^t)| \leq \sigma_a, \quad (10)$$

where  $c_1^i = (\mathbf{p}_1^s, \mathbf{p}_1^t)$  and  $c_2^i = (\mathbf{p}_2^s, \mathbf{p}_2^t)$ ;  $(\mathbf{L}_1^s, \mathbf{L}_1^t)$  and  $(\mathbf{L}_2^s, \mathbf{L}_2^t)$  are associated LRA pairs of  $c_1^i$  and  $c_2^i$ ;  $\sigma_d$  and  $\sigma_a$  represent the distance and angle thresholds, respectively. If current sampled correspondence pair does not satisfy the constraints in Eqs. (9) and (10), the algorithm switches to next iteration. Otherwise, the transformation  $\mathbf{T}_i$  that aligns the two oriented points in  $\mathcal{R}_s$  to the point pair in  $\mathcal{R}_t$  is computed.

Traditional methods, such as RANSAC, use the inlier count in  $\mathcal{C}$  to judge the correctness of  $\mathbf{T}_i$ . They may fail to return a correct transformation when  $\mathcal{C}$  is severely contaminated by outliers. Instead, we use the count of “overlapped” points  $\mathcal{R}^*$  between the transformed  $\mathcal{R}_s$  and  $\mathcal{R}_t$  as a new discriminant criterion, which is denoted by  $N_{s \wedge t}$ . In particular,  $\mathcal{R}^*$  is defined as:

$$\mathcal{R}^* = \{\mathbf{p}_n^s | \min_{\mathbf{p}_j^t \in \mathcal{R}_t} \|\mathbf{p}_n^s \cdot \mathbf{R}_i + \mathbf{t}_i - \mathbf{p}_j^t\| \leq d_o, \mathbf{p}_n^s \in \mathcal{R}_s\}, \quad (11)$$



**Fig. 7.** Feature correspondences and pairwise range image registration results of 6 sample objects in the experimental dataset. For clarity, the source range image, target range image and correspondence are hereinafter shown in red, green and cyan, respectively.

where  $\mathbf{R}_i$  and  $\mathbf{t}_i$ , respectively, represent the rotation matrix and translation vector in  $\mathbf{T}_i$ , and  $d_0$  is a distance threshold to judge whether a transformed point in  $\mathcal{R}_s$  is close to  $\mathcal{R}_t$ . By repeating the above steps, the transformation that yields to the maximum  $N_{s \wedge t}$  is taken as the optimum transformation  $\mathbf{T}^*$ . Algorithm 1 provides the pseudo-code of 2SAC-GC.

---

**Algorithm 1:** The 2SAC-GC Algorithm.

---

**Input:** The correspondence set  $\mathcal{C}$ , the source and target range images  $\mathcal{R}_s$  and  $\mathcal{R}_t$ ;  
**Output:** The optimal transformation  $\mathbf{T}^*$ ;  
 Initial the overlapped point count  $N_{s \wedge t}^0 = 0$ ;  
**for**  $i = 1; i \leq N_{iter}$  **do**  
   Randomly select two correspondences  $c_1^i$  and  $c_2^i$  from  $\mathcal{C}$ ;  
   **if**  $c_1^i$  and  $c_2^i$  satisfy Eqs. (9) and (10) **then**  
     Compute a plausible transformation  $\mathbf{T}_i$  for  $c_1^i$  and  $c_2^i$ ;  
     Calculate  $N_{s \wedge t}^i$  using Eq. (11);  
     **if**  $N_{s \wedge t}^i > N_{s \wedge t}^0$  **then**  
        $N_{s \wedge t}^0 \leftarrow N_{s \wedge t}^i$ ;  
        $\mathbf{T}^* \leftarrow \mathbf{T}_i$ ;  
    $i \leftarrow i + 1$ ;  
**Return**  $\mathbf{T}^*$ ;

---

### 3.4. Fine registration

The coarse registration of  $\mathcal{R}_s$  and  $\mathcal{R}_t$  is finished after launching the 2SAC-GC algorithm. At last, the result is further refined using a modified ICP algorithm [36] proposed recently, which exhibits efficient and accurate performance. This process iteratively minimizes the registration error of the coarse registration, and merges the input two range images into a seamless shape.

## 4. Experimental results

In this section, experiments including qualitative results (Section 4.1), quantitative results (Section 4.2), and comparisons with the state-of-the-arts (Section 4.3) on both Laser Scanner and Kinect datasets. The Laser Scanner dataset incorporates 32 pairs of high-resolution range images with 16 pairs from the Stanford 3D scanning repository [44], and 16 pairs from the UWA 3D Modeling dataset [50]. The Kinect dataset contains 51 pairs (a pair from each category) of low-resolution range images in the Washington RGB-D (Kinect) object dataset [51].

The parameter settings of the registration algorithm (if not specifically noted) based on the analysis in Section 3 are:  $k = 50$ ,  $\sigma_d = 3$  mr,  $\sigma_a = 10^\circ$ ,  $d_0 = 5$  mr, and  $N_{iter} = 200$  for the experiments

in this paper. All the experiments in this paper are implemented on a PC with an Intel Core i3-2120 3.3 GHz CPU and 8GB RAM.

### 4.1. Qualitative results

We adopt our pairwise registration method to generate point-to-point correspondences and align range images with different poses and partial overlaps, as shown in Fig. 7. Clearly, our method is capable of producing sufficient consistent correspondences between the source and target range images. Accordingly, precise registrations are achieved on range image pairs taken from both Laser Scanner and Kinect datasets. Note that the input range images are given in arbitrary initial positions and the registration course is fully automatic, which indicates no human interventions are participated.

We also present another 20 registration results from the Laser Scanner and Kinect datasets, as shown in Fig. 8. It is visually obvious that our method successfully aligns all the range image pairs in Fig. 8 when faced with rigid transformations, partial overlaps, noise (e.g., the Bell pepper and Greens), and featureless shapes (e.g., the Notebook and Coffee mug).

### 4.2. Quantitative results

In addition to the qualitative results, we also give quantitative results of all the range image pairs in the Laser Scanner and Kinect datasets. Particularly, the registration accuracy, robustness against noise and varying mesh resolutions, and computational costs are given below.

#### 4.2.1. Registration accuracy

Two criteria are employed for quantitative assessment of our registration method, i.e., the rotation error  $\varepsilon_r$  between the estimated rotation matrix  $\mathbf{R}_E$  and the ground truth one  $\mathbf{R}_{GT}$ , and the translation error  $\varepsilon_t$  between the estimated translation vector  $\mathbf{t}_E$  and the ground truth one  $\mathbf{t}_{GT}$  [13,37]. Specifically, they are defined as:

$$\varepsilon_r = \arccos\left(\frac{\text{trace}(\mathbf{R}_d) - 1}{2}\right) \frac{180}{\pi}, \quad (12)$$

$$\varepsilon_t = \frac{\|\mathbf{t}_{GT} - \mathbf{t}_E\|}{mr}, \quad (13)$$

where  $\mathbf{R}_d = \mathbf{R}_{GT}(\mathbf{R}_E)^{-1}$ . Here,  $\mathbf{R}_{GT}$  and  $\mathbf{t}_{GT}$  are obtained by first manual alignment and then ICP refinement. The registration results of the whole 83 pairs of range images are shown in Table 2.

One can see that, most range images have been accurately aligned (i.e., with small registration error). To judge the correctness of a pairwise registration, thresholds are required as principles. For example, we treat a pairwise registration as correct

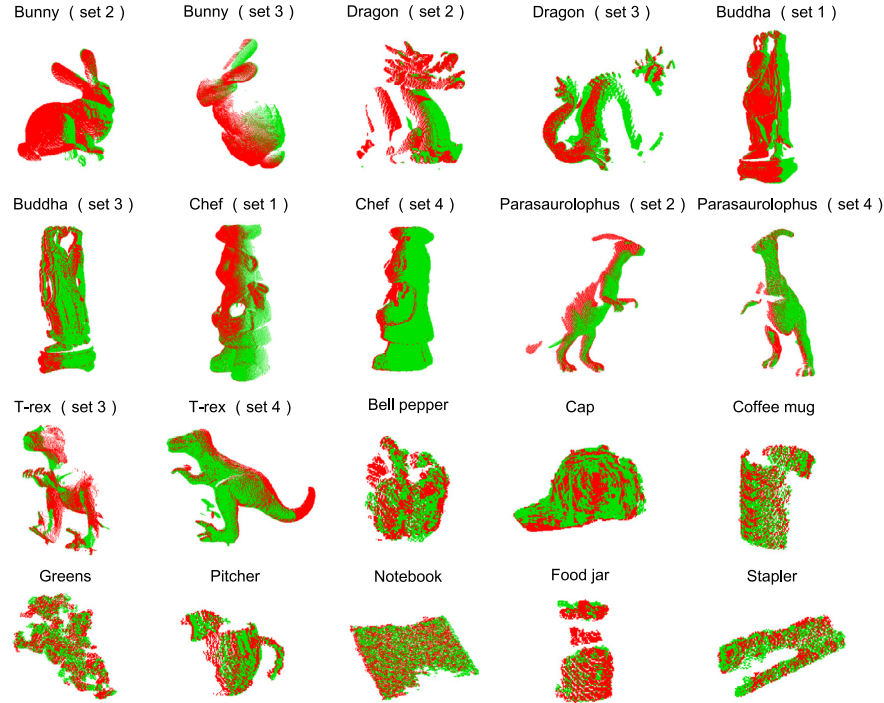


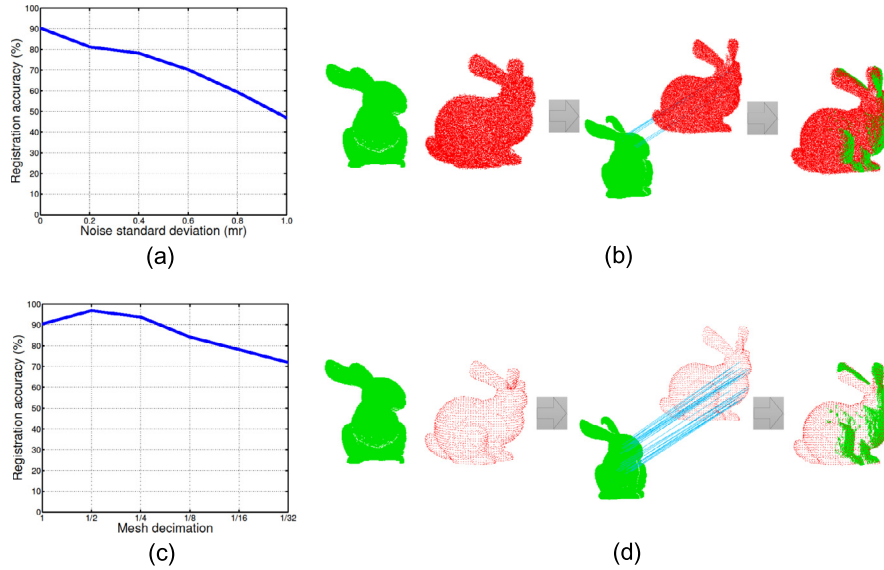
Fig. 8. Pairwise registration results of 20 pairs of range images in the experimental dataset.

Table 2

Registration accuracy of range image pairs in the Laser Scanner and Kinect datasets. Y indicates the registration is correct, and N denotes false registration. Specifically, a registration is judged as correct only if its rotation error and translation error are simultaneously smaller than  $6^\circ$  and 6 mr.

<i>(a) Laser scanner</i>											
Data	Armadillo (set 1)	Armadillo (set 2)	Armadillo (set 3)	Armadillo (set 4)	Budhha (set 1)	Budhha (set 2)	Budhha (set 3)	Budhha (set 4)	Bunny (set 1)	Bunny (set 2)	Bunny (set 3)
Error $\varepsilon_r$ ( $^\circ$ )	0.0000	0.6085	0.4370	0.7533	4.5342	0.7631	0.2324	0.8113	0.2934	3.8762	2.1260
Error $\varepsilon_t$ (mr)	1.0117	1.7116	1.3489	1.7454	2.3036	1.8312	0.4829	2.7489	0.8090	11.1123	7.4187
Result (Y/N)	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N
Data	Bunny (set 4)	Dragon (set 1)	Dragon (set 2)	Dragon (set 3)	Dragon (set 4)	Chef (set 1)	Chef (set 2)	Chef (set 3)	Chef (set 4)	Chicken (set 1)	Chicken (set 2)
Error $\varepsilon_r$ ( $^\circ$ )	0.5374	1.9984	0.8523	0.2533	0.5781	0.7285	2.2459	1.6599	0.8541	1.1114	135.75
Error $\varepsilon_t$ (mr)	0.5215	3.0928	5.6267	0.2103	1.5153	0.8007	5.5598	1.3043	1.4474	0.8704	77.9993
Result (Y/N)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Data	Chicken (set 3)	Chicken (set 4)	Parasaurolophus (set 1)	Parasaurolophus (set 2)	Parasaurolophus (set 3)	Parasaurolophus (set 4)	T-rex (set 1)	T-rex (set 2)	T-rex (set 3)	T-rex (set 4)	
Error $\varepsilon_r$ ( $^\circ$ )	1.2108	0.3090	1.5250	0.6501	0.7209	0.2094	1.3206	0.5012	1.0806	0.7039	
Error $\varepsilon_t$ (mr)	1.4269	0.2977	1.1060	1.2408	1.1346	0.0947	4.2946	1.1793	0.7230	0.7959	
Result (Y/N)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
<i>(b) Kinect</i>											
Data	Apple	Ball	Banana	Bell-pepper	Binder	Bowl	Calculator	Camera	Cap	Cell-phone	Cereal-box
Error $\varepsilon_r$ ( $^\circ$ )	2.4070	14.7863	1.4904	4.1912	0.5902	1.0141	2.5130	4.6094	2.1145	3.5835	1.2399
Error $\varepsilon_t$ (mr)	0.2392	0.9657	0.6442	0.7974	7.3875	0.6984	0.8364	2.2660	2.0409	2.5111	9.6418
Result (Y/N)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	N
Data	Coffee-mug	Comb	Dry_battery	Flashlight	Food_bag	Food_box	Food_can	Food_cup	Food_jar	Garlic	Glue-stick
Error $\varepsilon_r$ ( $^\circ$ )	2.7209	7.9671	2.5401	2.0734	0.3275	0.5679	4.5129	0.0791	2.2321	3.0766	3.9030
Error $\varepsilon_t$ (mr)	0.9521	0.6583	2.3716	0.4539	0.3772	0.3235	1.2308	1.1312	0.3580	0.4529	0.9023
Result (Y/N)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Data	Greens	Hand_towel	Instant_noodles	Keyboard	Kleenex	Lemon	Lightbulb	Lime	Marker	Mushroom	Notebook
Error $\varepsilon_r$ ( $^\circ$ )	0.0560	1.0598	0.9553	180.0005	179.4019	4.6333	3.7552	7.7365	4.1023	4.3602	1.2652
Error $\varepsilon_t$ (mr)	0.4601	0.8182	0.6826	12.6733	18.8777	0.3168	0.6776	2.0347	1.1482	0.7384	4.1851
Result (Y/N)	Y	Y	Y	N	N	Y	Y	N	Y	Y	Y
Data	Onion	Orange	Peach	Pear	Pitcher	Plate	Pliers	Potato	Rubber-eraser	Scissors	Shampoo
Error $\varepsilon_r$ ( $^\circ$ )	10.7756	5.4426	0.0000	0.0000	1.0900	0.5219	2.1344	4.5192	2.0959	1.0500	2.2379
Error $\varepsilon_t$ (mr)	1.5192	1.2575	0.7608	0.7630	0.3743	1.3731	0.5653	0.3458	4.0797	0.2400	6.3742
Result (Y/N)	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Data	Soda_can	Sponge	Stapler	Tomato	Tooth_brush	Tooth_paste	Water_bottle				
Error $\varepsilon_r$ ( $^\circ$ )	2.3933	2.3739	1.0705	27.3569	2.8988	2.1078	1.5035				
Error $\varepsilon_t$ (mr)	0.9652	0.5861	0.9138	1.1368	2.2614	0.3976	1.0899				
Result (Y/N)	Y	Y	Y	N	Y	Y	Y				





**Fig. 9.** Robustness evaluation of the proposed pairwise registration method. (a) Robustness against noise. (b) The registration result of Bunny (set 2) with 1.0 mr Gaussian noise. (c) Robustness against varying mesh resolutions. (d) The registration result of Bunny (set 2) with  $\frac{1}{32}$  mesh decimation.

**Table 3**

Registration results on the Laser Scanner and Kinect datasets using the original and a set of modified versions of the proposed MaSH-based method. MaSH: *ld*, MaSH: *nd*, MaSH: *sva*, MaSH-*w* and MaSH – RP denote MaSH feature with single local depth, normal deviation, surface variation angle, without histogram-weights *w* (in Eq. (5)), and without radial partition, respectively.

	Original	MaSH: <i>ld</i>	MaSH: <i>nd</i>	MaSH: <i>sva</i>	MaSH – <i>w</i>	MaSH–RP
Laser scanner (%)	90.36	53.13	56.25	65.63	81.25	71.88
Kinect (%)	80.39	60.78	54.90	58.82	72.55	58.82

only if its rotation error and translation error are simultaneously smaller than  $6^\circ$  and 6 mr. Based on current thresholds, our method achieves a registration accuracy of 90.36% and 80.39% on the Laser Scanner and Kinect datasets, respectively. Reasons to explain a drop of performance on the Kinect dataset are: the range images acquired by the Kinect sensor are quite noisy, and most objects in the Kinect dataset exhibit poor geometric features. In general, our method achieves high registration accuracy on both datasets.

To give an insightful analysis on our MaSH-based registration method, we also present the registration results (Table 3) on the experimental datasets with some modifications on our original method. As witnessed by the table, the deterioration of performance is rather notable for the methods using single attribute description. That is, a drop of 30–40% of registration accuracy on the Laser Scanner dataset, and 20–30% on the Kinect dataset can be found for the modified versions with single attribute description. The result neatly reveals the importance of feature representation from multiple aspects. Another observation is that when the weighting term *w* is removed from MaSH, the registration performance drops as well. It can be explained that without the weighting constrain, all the sub-histograms in MaSH have the same contribution for feature description, making the descriptor to be more sensitive to self-occlusion. We also test the registration performance when the spatial information in MaSH is discarded. One can see from the table that a clear drop of performance on both datasets is resulted, which highlights the importance of employing spatial information for feature encoding.

#### 4.2.2. Robustness to noise

In order to test the robustness of our registration method against noise, we add a Gaussian noise with increasing standard

deviation of 0.2 mr, 0.4 mr, 0.6 mr, 0.8 mr, and 1.0 mr to the target range images in the Laser Scanner dataset. The registration accuracy results with respect to different levels of noise are shown in Fig. 9(a).

We can find that the registration accuracy stays being higher than 70% when the standard deviation of noise is smaller than 0.6 mr. As the standard deviation gets larger, the registration accuracy drops quickly. It can be inferred that, the proposed registration method is robust to low and medium levels (i.e., the standard deviation is smaller than 0.6 mr) of noise, whereas turns to be sensible to high levels (i.e., the standard deviation is larger than 0.6 mr) of noise. It is because noise with large standard deviations has a strong impact on the calculation of normals [37], which are the basis of the calculation of two attributes (i.e., normal deviation and surface variation angle) in MaSH. Fig. 9(b) also shows the registration result of Bunny (set 2) with 1.0 mr noise. One can see that our method still can bring the source range image close to the noisy target range image, which demonstrates that our method has certain applicability even under high levels of noise.

#### 4.2.3. Robustness to varying mesh resolutions

We further evaluate the robustness of the proposed registration method to varying mesh resolutions. In this experiment, all the target range images are respectively downsampled to  $\frac{1}{2}$ ,  $\frac{1}{4}$ ,  $\frac{1}{8}$ ,  $\frac{1}{16}$ , and  $\frac{1}{32}$  of their original mesh resolutions. The results are shown in Fig. 9(c).

Clearly, our method is quite robust to mesh resolution variation. The registration performance with  $\frac{1}{8}$  is even comparable to the original, and still higher than 70% when the mesh decimation level reaches  $\frac{1}{32}$ . Fig. 9(d) presents the registration result of Bunny (set 2) with  $\frac{1}{32}$  mesh decimation. The points in the target range image are obviously quite sparse, while the points in the source one are fairly dense. We can see that our method still generates enough consistent correspondences, and achieves a success registration in such challenging case.

#### 4.2.4. Registration timing statistics

Table 4 reports the time costs of our registration method with respect to different data resolutions. Note that our code is implemented in Microsoft Visual C++ 2010 with no parallel computing techniques.

**Table 4**

Registration time costs with respect to different data resolutions. ( $T_{kd}$ ,  $T_{fe}$ ,  $T_{fm}$ ,  $T_{te}$ ,  $T_{rf}$  and  $T_{sum}$  represent the computation time of keypoint detection, feature extraction, feature matching, transformation estimation, refining and the whole registration course, respectively.)

Data	Source size	Target size	$T_{kd}$ (s)	$T_{fe}$ (s)	$T_{fm}$ (s)	$T_{te}$ (s)	$T_{rf}$ (s)	$T_{sum}$ (s)
Marker	540	594	0.001	0.063	0.005	1.279	0.108	1.456
Orange	2,640	2,536	0.005	0.553	0.024	11.193	0.605	12.380
Bowl	10,833	10,600	0.022	2.067	0.273	9.928	0.305	12.595
Armadillo (set 1)	24,029	28,220	0.061	6.090	2.505	15.386	6.586	30.628
Bunny (set 1)	40,097	40,256	0.089	9.504	2.835	25.092	10.431	47.951

**Table 5**

Parameter settings for eight feature descriptors.

	Support radius (mr)	Dimensionality	Length
Spin image	20	15 × 15	225
Snapshots	20	40 × 40	1600
THRIFT	20	32 × 1	32
FPFH	20	3 × 11	33
RoPS	20	3 × 3 × 3 × 5	135
LFSH	20	10 + 15 + 5	30
RCS	20	6 × 12	72
MaSH	20	5 × (10 + 8 + 10)	140

Two major observations can be made from the results. First, the main time costs of our registration method are spent on the transformation estimation and refining stages. It is owing to that both processes require iterations. Particularly, the time cost on transformation estimation (i.e., the designed 2SAC-GC algorithm) is particularly striking. It is owing to that all points in both range images participate in the transformation matrix estimation to achieve global optimum. It is therefore a trade-off between transformation estimation accuracy and calculation speed. Second, the time efficiency of our registration method is closely related to the resolution of the input range images. More specifically, range images with higher data resolution needs more time costs. In general, low-resolution range images (e.g., Marker and Orange) require a few seconds merely, and high-resolution range images (e.g., Armadillo (set 1)) that even contain tens thousands of points are aligned within one minute.

#### 4.3. Comparisons with the state-of-the-arts

In order to further validate the superiority of the our techniques, we first compare the proposed 2SAC-GC algorithm with the classical RANSAC algorithm [32] and its recent variant, i.e., the 1P-RANSAC algorithm [36]. Then, the proposed MaSH descriptor is compared with seven state-of-the-art descriptors, i.e., spin image [18], snapshots [24], THRIFT [15], FPFH [8], RoPS [25], LFSH [29], and RCS [26].

The implementation details of the evaluated methods are as follows. The compared RANSAC algorithm exists in the PCL library, and 1P-RANSAC algorithm is our re-implementation of [36] using C++. Regarding the feature descriptors, the compared spin image, FPFH and RoPS descriptors are just PCL functions (we use PCL 1.8.0 version). Note that the spin image descriptor actually has two versions at present. One is the original version suggested by the authors of spin image, i.e., using a cylindrical neighboring points. The other is a modified one capsuled in PCL, where spherical neighbors are used, and we take the PCL version in this paper to achieve scale consistency with other descriptors. Snapshots, THRIFT are our re-implementations by following the very details provided by the authors, while LFSH and RCS are implemented using the source C++ code provided by the authors. Details related to the parameters of these descriptors are summarized in Table 5.

##### 4.3.1. 2SAC-GC vs. the state-of-the-arts

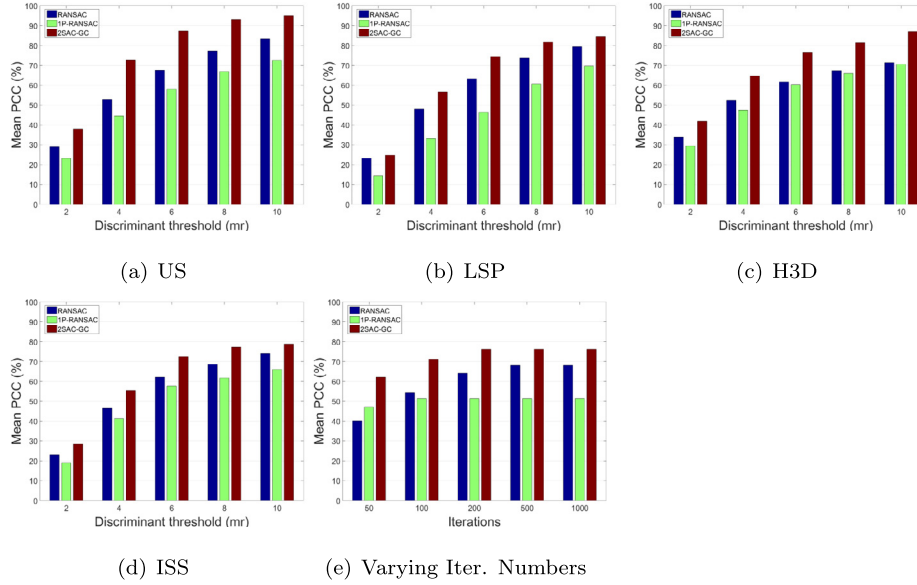
We use the percentage of correct correspondences (PCC) [37] to evaluate the performance of the 2SAC-GC and RANSAC algorithms. To judge the correctness of a correspondence, the following principle is employed:

$$\|(\mathbf{p}_i^s \cdot \mathbf{R}_{GT} + \mathbf{t}_{GT}) - \mathbf{p}_i^t\| \leq \varepsilon_c, \quad (14)$$

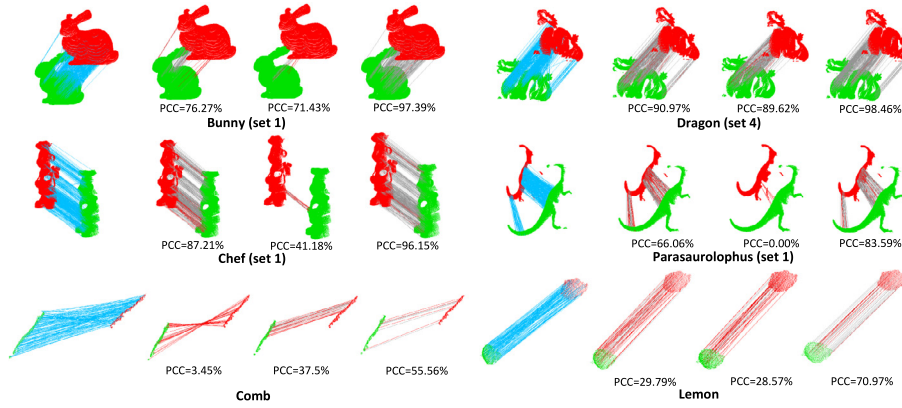
where  $\mathbf{p}_i^s$  and  $\mathbf{p}_i^t$  are two corresponding points in correspondence  $c_i$ , and  $\varepsilon_c$  is a tunable threshold. The smaller  $\varepsilon_c$ , the more accurate  $c_i$  is. We respectively use the RANSAC, 1P-RANSAC and 2SAC-GC algorithms to estimate a transformation for each range image pair in the Laser Scanner and Kinect datasets. In addition to uniform sampled (US) keypoints in the designed registration algorithm, we also test the performance of 2SAC-GC when using other standard 3D keypoint detectors including the local surface patches (LSP) [42], Harris 3D (H3D) [52] and intrinsic shape signature (ISS) [53]. To compare fairly, the same correspondence set (i.e., the initial  $k$  correspondences as described in Section 3.2) is taken as the input for all the tested algorithms. By varying the discriminant threshold  $\varepsilon_c$  and the number of iterations  $N_{iter}$ , the mean PCC results of all range image pairs are shown in Fig. 10.

It is clear that our 2SAC-GC algorithm outperforms the RANSAC and 1P-RANSAC algorithms at all levels of threshold when combined with different keypoint detectors (Fig. 10(a)–(d)) and iterations (Fig. 10(e)). It can be explained that both the RANSAC and 1P-RANSAC algorithms employ the inlier count of correspondences as the judging criterion, which can hardly deal with the case that more than 50% of the correspondences are false. On the contrary, the proposed 2SAC-GC instead judges the correctness of a pairwise registration from a global perspective, i.e., using the “overlapping” point count instead of the inlier count of correspondences. Meanwhile, the 1P-RANSAC algorithm is inferior to the RANSAC algorithm on the experimental datasets. It is due to that the 1P-RANSAC algorithm requires LRF calculation for each keypoint, and is therefore quite sensitive to LRF calculation error. Unfortunately, it is still a challenging task to guarantee a LRF’s robustness when noise, occlusion, point irregularity exist simultaneously. Note that the LRF method in the 1P-RANSAC algorithm is the one proposed in [25] as the authors suggest. Another observation is that although the 1P-RANSAC algorithm behaves poorly in terms of Mean PCC performance, it needs only 100 iterations to achieve the best performance, followed by our 2SAC-GC algorithm which needs 200 iterations as shown in Fig. 10(e). The RANSAC algorithm reaches the best performance with a cost of 500 iterations, which is owing to that it needs three correspondences at each iteration, in opposite to two correspondences in the 2SAC-GC algorithm and one in the 1P-RANSAC algorithm. The time costs of the RANSAC, 1P-RANSAC and 2SAC-GC algorithms with 100 iterations are 0.19 s, 0.14 s and 2.97 s, respectively. The proposed 2SAC-GC needs more time consumption for transformation estimation because it calculates the inliers for two whole range images rather than the correspondence set. Nevertheless, our 2SAC-GC achieves the best Mean PCC performance under all levels of discriminant thresholds and iterations.

We also present some visual results the three estimators with  $\varepsilon_c = 6$  mr and  $N_{iter} = 200$  using uniform sampled keypoints in



**Fig. 10.** Comparison of 2SAC-GC with RANSAC and 1P-RANSAC. From (a)–(d), the mean PCC results with respect to varying discriminant thresholds when respectively using the US (the keypoint detection method in our registration pipeline), LSP, H3D and ISS 3D keypoint detectors. (e) Shows the performance with respect to varying iteration numbers.



**Fig. 11.** Identified inliers of six different correspondence sets by RANSAC, 1P-RANSAC and 2SAC-GC with  $\varepsilon_c = 6$  mr and  $N_{iter} = 200$ . From left to right for each object, the initial correspondence set, the results of RANSAC, 1P-RANSAC and 2SAC-GC, respectively. For clarity, the cyan, gray and red lines, respectively, represent the initial, correct and false correspondences. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Fig. 11.** One can see that for relatively consistent input correspondences (e.g., correspondences in Bunny (set 1) and Dragon (set 4)), all the estimators achieve high PCC results. In particular, our 2SAC-GC performs slightly better in this case. For correspondence sets with large amount of outliers (e.g., the correspondences in Comb and Lemon), the RANSAC and 1P-RANSAC algorithms return particularly low PCC values, while the proposed 2SAC-GC still achieves high PCC values in such challenging case.

#### 4.3.2. MaSH vs. the state-of-the-arts

As for MaSH, we have deployed two kinds of experiments to examine its performance under range image registration context. Meanwhile, comparisons with several existing ones are also provided for better judgment. In the first experiment, we examine the performance of these feature descriptors within the proposed range image registration pipeline using the parameter settings in Table 5. The goal is to judge a feature descriptor's effectiveness in a whole registration algorithm. Table 6 shows the registration performance on the Laser Scanner and Kinect datasets when using different descriptors. The result reveals the superiority of our MaSH descriptor against the compared methods in terms of range image registration accuracy. Note that all the tested descriptors follows

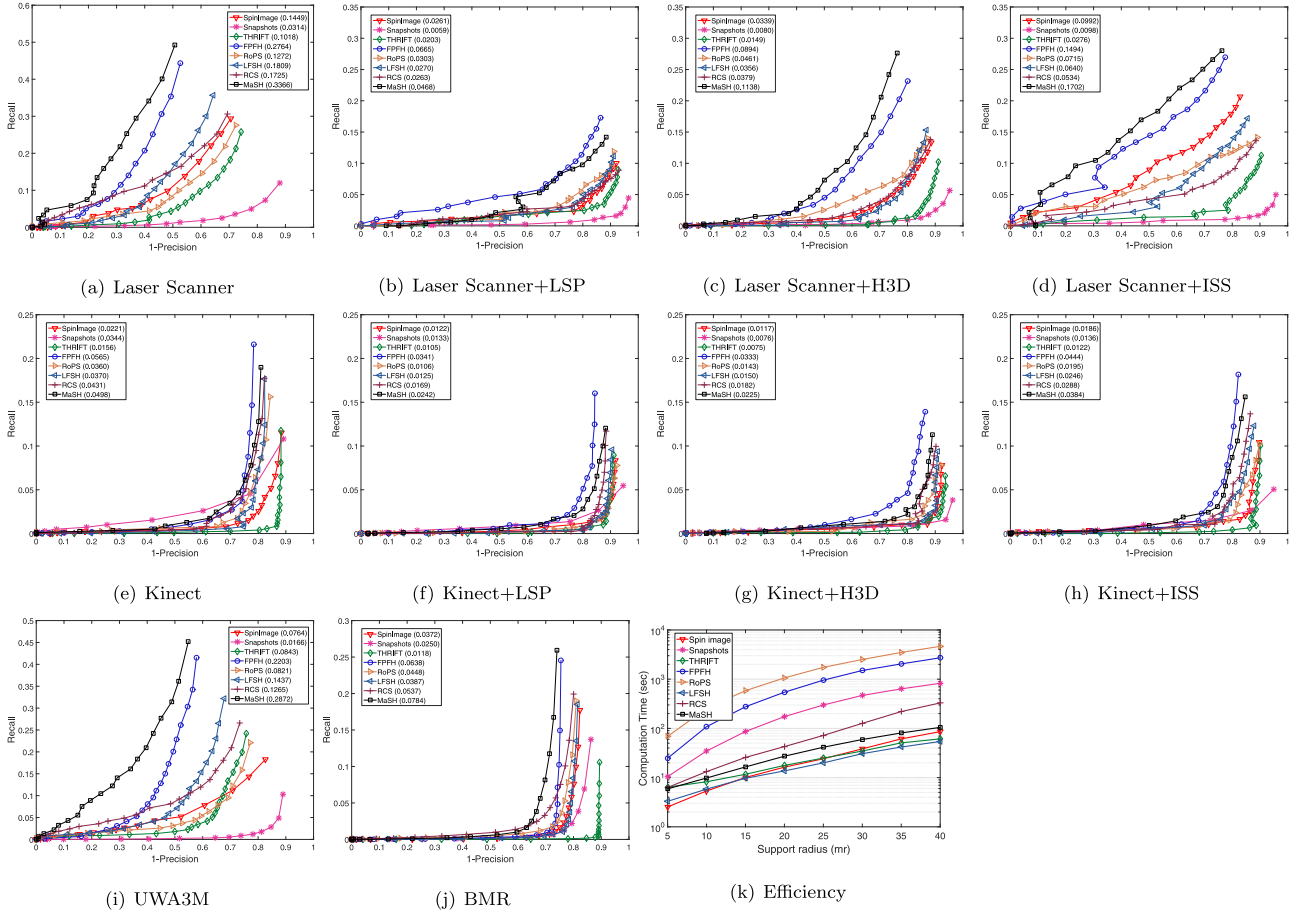
**Table 6**

Range image registration performance on the Laser Scanner and Kinect datasets using different feature descriptors. The best performance is marked with bold font.

	Spin image	Snapshots	THRIFT	FPFH	RoPS	LFSH	RCS	MaSH
Laser scanner (%)	81.25	53.13	68.75	87.50	56.25	75.00	78.12	<b>90.36</b>
Kinect (%)	60.78	31.37	54.90	64.71	70.59	54.90	68.63	<b>80.39</b>

the registration pipeline (Fig. 6), the only difference is that different feature descriptors are used to perform feature extraction.

In the second experiment, we test the feature matching performance of these feature descriptors independently, i.e., without the influence of keypoint detection, transformation estimation techniques capsuled in a whole registration algorithm. The quantitative performance of a feature descriptor is measured by recall versus 1-precision (RPC) [25,30,31] in this paper. Also, to provide more clarity for comparison, the area under the precision versus recall curve (PRC) AUC<sub>pr</sub> [31] measure is used. AUC<sub>pr</sub> is a simple and aggregated metric to measure how an algorithm performs over the whole precision–recall space [54]. Note that it is straightforward



**Fig. 12.** Feature matching performance of eight feature descriptors on range image registration datasets.  $AUC_{pr}$  values are shown in the parentheses in (a)–(j). (k) is the time efficiency result. For clarity, the Y-axis in (k) is the logarithmic axis.

to transform an RPC to PRC. To calculate an RPC, 1000 keypoints are randomly selected (if not specifically noted) from each source range image [31]. By calculating their corresponding points in the target range image using the ground truth transformation, the ground truth correspondences are generated. Then, based on the parameter settings in Table 5, spin image, snapshots, THRIFT, FPFH, RoPS and MaSH descriptors are extracted for the keypoints in the source and target range images. In addition to ground truth keypoints, we also justify the feature matching performance when using standard keypoints, i.e., the LSP, H3D and ISS detected keypoints on both the source and target range images. The purpose is to examine the feature matching performance of a feature descriptor with the interference of keypoint localization error. Follow the scheme in [20,25], we obtain the RPC and  $AUC_{pr}$  results on the Laser Scanner dataset and the Kinect dataset as shown in Fig. 12(a)–(h).

We can see that our MaSH descriptor achieves the best performance on the Laser Scanner dataset when using the ground truth, H3D and ISS keypoints, and ranks the second with the LSP keypoints. Also, one can find that the performance of all descriptors drops when undergoing keypoint localization error. For the Kinect dataset, which is more challenging for feature matching due to noise, holes and featureless shapes, the proposed MaSH descriptor ranks the second thorough all types of keypoints and is slightly inferior to the FPFH descriptor. In addition to current Laser Scanner and Kinect datasets, we further test the RPC performance of these descriptors on the UWA3M (the whole one) and Bologna Mesh Registration (BMR) [30] datasets. The former is acquired using a LiDAR scanner and the latter is obtained by Kinect, they respectively

contain 496 and 485 valid (i.e., with at least 10% overlap ratio) range image pairs. The results are respectively shown in Fig. 12(i) and (j). In these two large-scale datasets, our MaSH descriptor consistently surpasses all the compared ones, showing the advantage of our MaSH descriptor in the context of range image registration. The results suggest that the MaSH descriptor could resist the nuisances exist in range image registration scenario quite well.

The superiority of MaSH in terms of feature matching can be explained by at least three reasons. First, MaSH performs a comprehensive description for the local surface via the integration of multiple geometric attributes, in contrast to the methods with single attribute description (e.g., spin image and snapshots). Second, MaSH consists of both geometric and spatial information encodings and therefore turns to be more distinctive than the ones with only geometric information encoding (e.g., THRIFT and LFSH). Third, the sub-histograms in the MaSH descriptor are attached with weights relating to their spatial locations, in order to keep robust to incomplete boarders and self-occlusion.

Finally, we present the time efficiency results of the eight descriptors in Fig. 12(k). The curves are obtained as follows. First, 1000 keypoints are randomly selected from the source range images in the turning dataset (Fig. 5). Then, feature descriptors with varying support radii are calculated for these keypoints. Finally, the total time costs of each descriptor for all the keypoints against varying support radii are recorded, therefore generating a curve. We note that the point count instead of the quality of the local surface is the major fact for the calculation of feature descriptors [31]. This experiment has no strict demands on the quality of datasets. From the curves in Fig. 12(k), we can see that the



proposed MaSH descriptor ranks medium among these descriptors, being surpassed by the spin image, LFSH and THRIFT descriptors. However, our MaSH descriptor outperforms these three ones by a large margin in terms of feature matching performance.

These results show that none of the compared descriptors manage to simultaneously achieve a pleasurable and balanced performance in terms of descriptiveness, robustness, time efficiency. For instance, FPFH achieves high RPC performance on both the Laser Scanner and Kinect datasets, though, it is very slow for feature extraction. THRIFT shows its superiority in terms of time efficiency, while it exhibits limited descriptiveness and robustness. Overall, the proposed MaSH descriptor attains to perform effective, robust, and efficient shape description for the local surface simultaneously.

Regarding the outcomes of this paper, several points should be explained here: (1) we have compared many state-of-the-art methods in this paper. Regarding recent proposals such as SHOT and RoPS, indeed, they achieved very good performance in their evaluated datasets, e.g., the Bologna Object Recognition [30], and UWA Object Recognition [50] datasets. However, they are under different application contexts, i.e., mainly 3D object recognition. Meanwhile, many of their designing techniques are related to their application scenarios. In contrast, all our evaluated datasets are range image registration datasets, indicating that different nuisances are contained, i.e., missing boarder regions, holes and self-occlusion. The compared methods has never been tested on our experimental datasets before. (2) In terms of range image registration, the prior evaluation work in [31] also shows that different descriptors behave quite different on across-dataset experiments. For examples, RoPS and SHOT achieve top-ranked performance in terms of object recognition, while they just behave moderately in 2.5D view matching case. (3) Our MaSH descriptor is particularly designed for range image registration rather than shape retrieval (with complete local shapes) and 3D object recognition (with cluttered and occluded local shapes). It is somehow not surprising that our MaSH behaves pleasurable under range image registration context, because all its designing principles including choosing LRA instead of LRF, histogram weights and partition along radial direction are particularly for robust range image registration.

## 5. Conclusions and future work

In this paper, we proposed (1) a MaSH descriptor, (2) a 2SAC-GC algorithm, and (3) a coarse-to-fine pairwise registration method for range images. Range image registration experiments on both Laser Scanner and Kinect datasets demonstrated that our registration method is accurate, efficient, and robust to noise and varying mesh resolutions. We also compared the proposed 2SAC-GC algorithm and MaSH descriptor with the state-of-the-arts. The superior results of our techniques further confirm that: (i) the proposed 2SAC-GC algorithm is cable of estimating a correct transformation from correspondences with numerous outliers; (ii) the MaSH descriptor is highly descriptive, robust to common nuisances including noise, mesh resolution variation and keypoint localization error. MaSH is efficient for calculation as well.

Our future work includes two interesting directions. One is to enhance the MaSH's flexibility towards RGB-D objects by integrating photometric attributes. The other is to extend the proposed MaSH-based pairwise registration method to multi-view registration method, so as to automatically build a complete model or scene from a set of unordered range images.

## Acknowledgment

The authors would like to acknowledge the Stanford 3D Scanning Repository, the University of Western Australia and the University of Washington for making their datasets available

to us. This work is supported by the National High Technology Research and Development Program of China (863 Program) under Grant 2015AA015904, the 2015 annual foundation of China Academy of Space Technology (CAST), the Guangdong Innovative and Entrepreneurial Research Team Program under Grant 2014ZT05G304, and China Postdoctoral Science Foundation under Grant 2014M562028.

## References

- [1] S.A.A. Shah, M. Bennamoun, F. Boussaid, A novel feature representation for automatic 3D object recognition in cluttered scenes, *Neurocomputing* 205 (2016) 1–15.
- [2] Y. He, Y. Mei, An efficient registration algorithm based on spin image for Lidar 3D point cloud models, *Neurocomputing* 151 (2015) 354–363.
- [3] Y. Guo, F. Soheli, M. Bennamoun, J. Wan, M. Lu, An accurate and robust range image registration algorithm for 3D object modeling, *IEEE Trans. Multimed.* 16 (5) (2014) 1377–1390.
- [4] F. Tombari, N. Fioraio, T. Cavallari, S. Salti, A. Petrelli, L. Di Stefano, Automatic detection of pole-like structures in 3D urban environments, in: *Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2014, pp. 4922–4929.
- [5] R.B. Rusu, Z.C. Marton, N. Blodow, M. Beetz, Persistent point feature histograms for 3D point clouds, in: *Proceedings of the Tenth International Conference on Intelligent Autonomous Systems*, 2008, pp. 119–128.
- [6] A.E. Johnson, M. Hebert, Surface matching for object recognition in complex three-dimensional scenes, *Image Vis. Comput.* 16 (9) (1998) 635–651.
- [7] R.B. Rusu, N. Blodow, Z.C. Marton, M. Beetz, Aligning point cloud views using persistent feature histograms, in: *Proceedings of the Twenty-first IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008, pp. 3384–3391.
- [8] R.B. Rusu, N. Blodow, M. Beetz, Fast point feature histograms (FPFH) for 3D registration, in: *Proceedings of the 2009 IEEE International Conference on Robotics and Automation*, 2009, pp. 3212–3217.
- [9] P.J. Besl, N.D. McKay, Method for registration of 3-D shapes, *IEEE Trans. Pattern Anal. Mach. Intell.* 14 (2) (1992) 239–256.
- [10] S. Rusinkiewicz, M. Levoy, Efficient variants of the ICP algorithm, in: *Proceedings of the third International Conference on 3-D Digital Imaging and Modeling*, 2001, pp. 145–152.
- [11] S. Ying, J. Peng, S. Du, H. Qiao, A scale stretch method based on ICP for 3D data registration, *IEEE Trans. Autom. Sci. Eng.* 6 (3) (2009) 559–565.
- [12] J. Salvi, C. Matabosch, D. Fofi, J. Forest, A review of recent range image registration methods with accuracy evaluation, *Image Vis. Comput.* 25 (5) (2007) 578–596.
- [13] A.S. Mian, M. Bennamoun, R.A. Owens, A novel representation and feature matching algorithm for automatic pairwise registration of range images, *Int. J. Comput. Vis.* 66 (1) (2006) 19–40.
- [14] A. Makadia, A. Patterson, K. Daniilidis, Fully automatic registration of 3D point clouds, in: *Proceedings of the 2006 IEEE Computer Vision and Pattern Recognition*, 1, 2006, pp. 1297–1304.
- [15] A. Flint, A.R. Dick, A. Van Den Hengel, Thrift: local 3D structure recognition, in: *Proceedings of the Ninth Conference on Digital Image Computing: Techniques and Applications*, 7, 2007, pp. 182–188.
- [16] U. Castellani, M. Cristani, S. Fantoni, V. Murino, Sparse points matching by combining 3D mesh saliency with statistical descriptors, in: *Computer Graphics Forum*, 27, Wiley Online Library, 2008, pp. 643–652.
- [17] J. Novatnack, K. Nishino, Scale-dependent/invariant local 3D shape descriptors for fully automatic registration of multiple sets of range images, in: *Proceedings of the 2008 European Conference on Computer Vision*, 2008, pp. 440–453.
- [18] A.E. Johnson, M. Hebert, Using spin images for efficient object recognition in cluttered 3D scenes, *IEEE Trans. Pattern Anal. Mach. Intell.* 21 (5) (1999) 433–449.
- [19] S. Ruiz-Correa, L.G. Shapiro, M. Melia, A new signature-based method for efficient 3-D object recognition, in: *Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition*, 1, 2001, pp. 1–769.
- [20] Y. Guo, F. Soheli, M. Bennamoun, et al., A novel local surface feature for 3D object recognition under clutter and occlusion, *Inf. Sci.* 293 (2015) 196–213.
- [21] A. Frome, D. Huber, R. Kolluri, T. Bülow, J. Malik, Recognizing objects in range data using regional point descriptors, in: *Proceedings of the 2004 European Conference on Computer Vision*, 2004, pp. 224–237.
- [22] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (4) (2002) 509–522.
- [23] A. Flint, A. Dick, A. Van den Hengel, Local 3D structure recognition in range images, *IET Comput. Vis.* 2 (4) (2008) 208–217.
- [24] S. Malassiotis, M.G. Strintzis, Snapshots: a novel local surface descriptor and matching algorithm for robust 3D surface alignment, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (7) (2007) 1285–1290.
- [25] Y. Guo, F. Soheli, M. Bennamoun, M. Lu, J. Wan, Rotational projection statistics for 3D local surface description and object recognition, *Int. J. Comput. Vis.* 105 (1) (2013) 63–86.
- [26] J. Yang, Q. Zhang, K. Xian, Y. Xiao, Z. Cao, Rotational contour signatures for robust local surface description, in: *Proceedings of the 2016 IEEE International Conference on Image Processing*, IEEE, 2016, pp. 3598–3602.

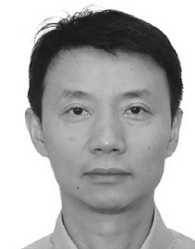
- [27] S.A.A. Shah, M. Bennamoun, F. Boussaid, A.A. El-Sallam, 3D-div: a novel local surface descriptor for feature matching and pairwise range image registration, in: Proceedings of the 2013 IEEE International Conference on Image Processing, IEEE, 2013, pp. 2934–2938.
- [28] S.A.A. Shah, M. Bennamoun, F. Boussaid, A novel 3D vorticity based approach for automatic registration of low resolution range images, *Pattern Recognit.* 48 (9) (2015) 2859–2871.
- [29] J. Yang, Z. Cao, Q. Zhang, A fast and robust local descriptor for 3D point cloud registration, *Inf. Sci.* 346 (2016) 163–179.
- [30] F. Tombari, S. Salti, L. Di Stefano, Unique signatures of histograms for local surface description, in: Proceedings of the 2010 European Conference on Computer Vision, 2010, pp. 356–369.
- [31] Y. Guo, M. Bennamoun, F. Soheli, M. Lu, J. Wan, N.M. Kwok, A comprehensive performance evaluation of 3D local feature descriptors, *Int. J. Comput. Vis.* 116 (1) (2016) 66–89.
- [32] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Commun. ACM* 24 (6) (1981) 381–395.
- [33] F. Tombari, L. Di Stefano, Object recognition in 3D scenes with occlusions and clutter by hough voting, in: Proceedings of the Fourth Pacific-Rim Symposium on Image and Video Technology (PSIVT), IEEE, 2010, pp. 349–355.
- [34] D. Thomas, A. Sugimoto, Robustly registering range images using local distribution of albedo, *Comput. Vis. Image Underst.* 115 (5) (2011) 649–667.
- [35] A.G. Buch, Y. Yang, N. Krüger, H.G. Petersen, In search of inliers: 3D correspondence by local and global voting, in: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2014, pp. 2075–2082.
- [36] Y. Guo, M. Bennamoun, F. Soheli, M. Lu, J. Wan, An integrated framework for 3-D modeling, object detection, and pose estimation from point-clouds, *IEEE Trans. Instrum. Meas.* 64 (3) (2015) 683–693.
- [37] Y. Guo, M. Bennamoun, F. Soheli, M. Lu, J. Wan, 3D object recognition in cluttered scenes with local surface features: a survey, *IEEE Trans. Pattern Anal. Mach. Intell.* 36 (11) (2014) 2270–2287.
- [38] A. Petrelli, L.D. Stefano, On the repeatability of the local reference frame for partial shape matching, in: Proceedings of the 2011 IEEE International Conference on Computer Vision, 2011, pp. 2244–2251.
- [39] A. Petrelli, L.D. Stefano, A repeatable and efficient canonical reference for surface matching, in: Proceedings of the 2012 IEEE Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 2012, pp. 403–410.
- [40] R. Bro, E. Acar, T.G. Kolda, Resolving the sign ambiguity in the singular value decomposition, *J. Chemom.* 22 (2) (2008) 135–140.
- [41] F. Phillips, E. O'Donnell, N. Kernis, Visual and haptic geometry of 3D shape discrimination, *J. Vis.* 15 (12) (2015) 866.
- [42] H. Chen, B. Bhanu, 3D free-form object recognition in range images using local surface patches, *Pattern Recognit. Lett.* 28 (10) (2007) 1252–1262.
- [43] M. Pauly, Point Primitives for Interactive Modeling and Processing of 3D Geometry, Citeseer, 2003 Ph.D. thesis.
- [44] B. Curless, M. Levoy, A volumetric method for building complex models from range images, in: Proceedings of the Twenty-third Annual Conference on Computer Graphics and Interactive Techniques, 1996, pp. 303–312.
- [45] M. Pauly, R. Keiser, M. Gross, Multi-scale feature extraction on point-sampled surfaces, in: *Computer Graphics Forum*, 22, Wiley Online Library, 2003, pp. 281–289.
- [46] A. Mian, M. Bennamoun, R. Owens, On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes, *Int. J. Comput. Vis.* 89 (2–3) (2010) 348–361.
- [47] A. Zaharescu, E. Boyer, R. Horaud, Keypoints and local descriptors of scalar functions on 2D manifolds, *Int. J. Comput. Vis.* 100 (1) (2012) 78–98.
- [48] J. Kammerl, N. Blodow, R.B. Rusu, S. Gedikli, M. Beetz, E. Steinbach, Real-time compression of point cloud streams, in: Proceedings of the 2012 IEEE International Conference on Robotics and Automation, 2012, pp. 778–785.
- [49] C. Papazov, D. Burschka, An efficient RANSAC for 3D object recognition in noisy and occluded scenes, in: Proceedings of the Tenth Asian Conference on Computer Vision (ACCV), Springer, 2010, pp. 135–148.
- [50] A.S. Mian, M. Bennamoun, R. Owens, Three-dimensional model-based object recognition and segmentation in cluttered scenes, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (10) (2006) 1584–1601.
- [51] K. Lai, L. Bo, X. Ren, D. Fox, A large-scale hierarchical multi-view RGB-D object dataset, in: Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2011, pp. 1817–1824.
- [52] I. Sipiran, B. Bustos, Harris 3D: a robust extension of the harris operator for interest point detection on 3D meshes, *Vis. Comput.* 27 (11) (2011) 963–976.
- [53] Y. Zhong, Intrinsic shape signatures: a shape descriptor for 3D object recognition, in: Proceedings of the Twelfth International Conference on Computer Vision Workshops (ICCV Workshops), IEEE, 2009, pp. 689–696.
- [54] J. Davis, M. Goadrich, The relationship between precision–recall and ROC curves, in: Proceedings of the Twenty-third International Conference on Machine Learning, ACM, 2006, pp. 233–240.



**Jiaqi Yang** received his B.S. degree from Huazhong University of Science and Technology, China in 2014. He is currently working toward the Ph.D. degree at the School of Automation, Huazhong University of Science and Technology. His research interests include 3D local shape description, 3D registration, 3D modeling and pattern recognition.



**Qian Zhang** received the B.E. and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2008 and 2013, respectively. He is currently working toward postdoctoral research at the School of Automation, Huazhong University of Science and Technology, Wuhan. His research interests include image registration, image fusion, and pattern recognition in remote sensing images.



**Zhiguo Cao** is currently a professor and Dean of the School of Automation, Huazhong University of Science and Technology. He received the B.S. degree in Electronics Engineering in 1985 and M.S. degree in Communication and Information System in 1990 from University of Electronics Science and Technology of China, and his Ph.D. degree in pattern recognition and intelligence systems in 2001 from Huazhong University of Science and Technology, China. His current research interests include Machine Learning and Pattern Recognition, Feature Representation and Extraction, Image Processing, etc.