Computers & Graphics (2021)

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Dual spin-image: A bi-directional spin-image variant using multi-scale radii for 3D local shape description

Daryl L. Bibissi[a], Jiaqi Yang[b,a,c,*], Siwen Quan[d], Yanning Zhang[a,c]

[a]School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China
[b]Ningbo Institute of Northwestern Polytechnical University, 218 Qingyi Road, Ningbo, 315103, China
[c]The National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, Xi'an 710129, China
[d]School of Electronics and Control, Chang'an University, Xi'an, 710064, China

ARTICLE INFO

Article history:
Received November 3, 2021

Keywords: Local shape descriptor, 3D point cloud, bi-direction, feature matching

ABSTRACT

Obtaining good feature descriptors for 3D local shape under different conditions, such as noise, clutter, occlusion, and limited overlap, is still a challenging task in computer vision. In this paper, we present a variant of spin-image called dual spin-image (Dual-SI) for 3D local shape description that encodes bi-directional information between an oriented point and its surrounding neighbors within a given radius. By accumulating the parameters that encode the 3D point information into a 2D space bidirectionally, we can generate two different signatures. These signatures are then concatenated to form the Dual-SI feature descriptor. Finally, we propose a multi-scale radius approach to address the problem of occlusion and make use of a weighted kernel to address the noise problem. We tested our Dual-SI feature descriptor on popular datasets addressing 3D object registration, 3D object recognition, and shape retrieval scenarios. We also conduct experiments for 3D point cloud registration to further evaluate the effectiveness of our method. Consensus experimental results show that our Dual-SI achieves top-ranked performance on datasets with various nuisances and application contexts.

1. Introduction

Computer vision has been a highly researched sector in information technology for many decades. Object registration and recognition are two hot research areas within the computer vision community. The main task of 3D object recognition is to determine 3D information such as the orientation, shape, position, or volume of an object from a cluttered scene. 3D object registration aims to estimate the rigid transformation between two point clouds. Both tasks have a tremendous number of applications in robotics, remote sensing, entertainment, biometric system, automation, medical image analysis, navigation system, reverse engineering, to name just a few [1]. In recent years, the rapid technological development and low-cost of hardware devices for 3D imaging sensors, such as the Microsoft Kinect and Intel RealSense, have made point cloud data more accessible to the research community. The availability of data with the progress of computational power has led to the development of many algorithms over the years. Both 3D object recognition and point cloud registration can be achieved when consistent point-to-point correspondences are established between two 3D point clouds, which can be viewed as a correspondence establishment problem.

Point cloud correspondence establishment relies on crafting good descriptors that fully encode the geometric and spatial information of a 3D shape. Point cloud descriptors can be classified as *global* and *local*. Global descriptors encode information about the entire object geometry. Global descriptors achieve

*Corresponding author: Tel.: +86-131-3622-8315;
e-mail: jqyang@nwpu.edu.cn (Jiaqi Yang), jqyang@nwpu.edu.cn (Jiaqi Yang)

good results for object recognition and classification, geometry analysis, and pose estimation [2, 3, 4]. Typical global descriptors algorithms include the Viewpoint Feature Histogram (VFH) [4], Clustered Viewpoint Feature Histogram (CVFH) [5], the Ensemble of Shape Function (ESF) [6], and many others. However, global descriptors present some drawbacks. Because they directly encode the entire raw data which may contain clutter, additional preprocessing steps such as segmentation are required to identify potential candidates. In addition, they are sensitive to occlusion and limited overlap. By contrast, local descriptors which are the main fundamental research area for 3D object recognition and registration show many advantages, e.g., the robustness to clutter, occlusion, and missing regions [1], and have been successfully applied to a number of 3D vision tasks such as 3D registration, face recognition, and shape retrieval [7, 8, 9]. Different from global descriptors, local descriptors focus on encoding the information of the geometry in the vicinity of a point into a feature vector representation [10].

Several local descriptors have been proposed over the past three decades, such as Point Signature [11], Exponential Map (EM) [12], Rotational Contour Signatures (RCS) [13], 3D Shape Context (3DSC) [14], Point Feature Histogram (PFH) [15], and Spin-image [16, 17]. These descriptors can be classified into two categories according to whether a local reference frame (LRF) is used or not. LRF-based ones first build an LRF in the local surface area and then extract features along each LRF-aligned axis. LRF-independent local descriptors mainly rely on the statistics of invariant point attributes such as normals and densities. We refer the readers to [1] for a comprehensive survey. A good feature descriptor should be highly descriptive and robust to a set of nuisances such as noise, clutter, occlusion, and missing regions. As verified in Sec. 4.3, most of existing feature descriptors still fail to achieve a balanced performance when confronted with different nuisances.

Motivated by these considerations, we propose a robust descriptor called dual spin-image (Dual-SI) , which is a bi-directional and multi-scale radius variant of spin-image. First, we interchangeably select an orientated point p ("orientated" indicating normal attached) centered at a defined region within a radius r, and an oriented point q in the neighborhood region of p to generate two spin-images. For clarity, we also refer to the point p in the oriented point pair (p, q) as the center point in the rest of the paper. The two spin-images are then concatenated to form a whole feature descriptor. Furthermore, we use a multi-scale strategy to alleviate the problem of occlusion, and leverage a kernel to enhance the robustness of the descriptor with respect to noise. Besides the bi-directional aspect of our Dual-SI feature descriptor, our multi-scale radius differs from [18] in that: 1) we generate only one dual spin-image pyramid, the bin size of each element of the pyramid is reduced or increased with the size of its corresponding dual spin-image; 2) each Dual-SI in the pyramid is weighted by a α value to achieve the robustness to occlusion; 3) all the signatures in the pyramid are merged to form the final Dual-SI (Fig. 8). The performance of the Dual-SI features are comprehensively evaluated on public datasets, i.e., the Bologna Retrieval [19], the Bologna Dataset5 [10], the Bologna Mesh Registration [10], the UWA 3D Object Recognition [20, 21], and the UWA 3D Modeling [22]. The experimental results show that our proposed Dual-SI feature consistently achieves top-ranked performance on different datasets with various nuisances when compared with state-of-the-arts. In summary, our paper has the following contributions:

- A new variant of the spin-image descriptor called Dual-SI, which encodes bi-directional information between a basis point and its neighbors.
- A multi-scale radius approach to enhance the robustness of Dual-SI to occlusion.
- A weighted kernel that smooths the generated dual spin-images by removing sharp details and noise in order to alleviate the noise problem.

The rest of this paper is arranged in the following order. Sec. 2 gives a literature review of the existing descriptors with an emphasis on the spin-image. Sec. 3 introduces technical details about the proposed Dual-SI descriptor. Sec. 4 shows the experimental results of the Dual-SI method on the several standard datasets. Sec. 5 draws the conclusions and presents potential future work.

2. Related work

Over the years there have been a number of algorithms proposed for 3D local surface feature descriptors. Because our work is a variant of Johnson and Hebert's spin-image, we will first review the existing spin-image approach [16, 17, 23], and later introduce other existing local shape descriptors.

2.1. Spin-image

Spin-image [16, 17, 24] is one of the most popular local surface feature descriptor. It is generated for an oriented point on a surface of an object (or a vertex of a mesh that represents the surface of an object) that is associated with a direction. The oriented point is defined in a plane $(p, \vec{v_p})$, where p is the position of the orientated point and $\vec{v_p}$ is the surface normal associated with p. The coordinates of the plane are given by α, i.e., the perpendicular distance to the line through p, and β, i.e., the line parallel to p from the oriented point tangent plane. Using this coordinate system, a spin-map M_O is defined to encode the 3D coordinate of a neighbor point q into the 2D space of a $(p, \vec{v_p})$ coordinate system. The spin-map function $M_O : R^3 \rightarrow R^2$ is defined as: $M_O(x) \rightarrow (\alpha, \beta)$

$$\alpha = \sqrt{\|q_i - p\|^2 - (\vec{v_p}.(q_i - p))^2}, \qquad \beta = \vec{v_p}.(q_i - p) \quad (1)$$

The spin-map is then applied to all the vertices of the surface mesh of an object. The results are binned into a 2D feature histogram, forming the spin-image where pixel values are bins that contains several 3D points encoded in a i, j location. For every point falling in the spin-image plane (α, β), bi-linear interpolation is used to spread the 2D points to four surrounding bins. Another important element in spin-image generation is

the resolution of the image. The width W_s and height H_s of the spin-image are given by:

$$H_s = \frac{2\beta_{max}}{b} + 1\ , \qquad W_s = \frac{\alpha_{max}}{b} + 1 \tag{2}$$

where β_{max}, α_{max} represent the maximum coordinate values of the spin-image plane, and b is the number of partitions along each axis of the spin-image coordinate.

To perform object recognition, the spin-images of vertices in the scene are compared to the spin-images of vertices in the model. A correspondence is established when a suitable match is identified. Spin-images of two corresponding points are similar if they are uniformly sampled. When they are not, Johnson and Hebert [24] employed a mesh simplification algorithm to achieve uniform sampling. Spin-images generated from the scene and the model are similar if they are derived from the shape of a common object. To measure the similarity between two spin-images, Johnson and Hebert [16] employed the confidence in the linear correlation coefficient given by:

$$C(P, Q) = (atanh(R(P, Q)))^2 - \lambda\left(\frac{1}{N-3}\right). \tag{3}$$

In the above equation, R is the correlation coefficient, P and Q the two spin-images, λ is the weight applied to the variance, and N is the number of bins that contain the same number of points in both P and Q. The term $R(P, Q)$ is given by:

$$R(P, Q) = \frac{n\sum PQ - \sum P \sum Q}{((N\sum P^2 - (\sum P)^2)(N\sum Q^2 - (\sum Q)^2))^{\frac{1}{2}}} \tag{4}$$

Many spin-image variants have been proposed to improve its distinctiveness and the robustness to different nuisances. For instance, multi-resolution spin-images [18] address the costly factor of the original spin-image when comparing two uncompressed spin-images and gives a better method to select the appropriate bin and image width for spin-images. Ruiz et al. [25] introduced an approach based on a set of discriminative descriptors called spherical spin-images, which encode information dispatched by classes of distributions of surface points from an object. Darom and keller [26] presented scale-invariant spin-image which is a scale-invariant version of the spin-image. Johnson and Hebert [24] introduced a variant of the spin-image that uses spherical parameterization. More recently, Lu et al. [27] proposed longitude and latitude spin-image (LLSI) that incorporates a local reference frame (LRF) module and a feature descriptor obtained by joining the longitude and latitude images to the original spin-image. Guo et al. [28] introduced Tri-spin-image (TRISI) using a local reference frame (LRF) to generate three different signatures by respectively projecting the neighboring points on the x-axis, y-axis and z-axis of the LRF, they further compressed the obtained signatures using the principal component analysis method. Liang et al. [29] proposed Spin Contour, which encodes a 2D point set containing boundary points of the image of the input local surface obtained by spin-map transformation. Spin Contour is shown to be robust to mesh resolution variation and noise.

2.2. Other local descriptors for object matching and recognition

Many descriptors have been proposed in the literature that encode local surface information relative to a surface point for recognition and registration purposes. They can be divided into two categories according to if the object is rigid or non-rigid. The main characteristic of deformable shape descriptor is to keep invariant to non-rigid transformation. Because our work is mainly related to the latter category (i.e., rigid shape), we suggest the readers referring to [30] for more details on the subject. Furthermore, local shape descriptors for rigid data can be classified as LRF-based and LRF-independent.

For LRF-independent descriptors, Point Feature Histogram (PFH) [15] computes a coordinate system from the normals of each pair of points on a local surface. The system is then used to compute a statistical histogram of the differences between the normals and the euclidean distance between the point pair. The Fast Point Feature Histogram (FPFH) [31] computes a statistical histogram of the differences between normals following PFH, but instead of using all points in the vicinity of a local surface, FPFH uses a direct connection between the key point and its neighbors, thus reducing the computational complexity from $O(nk^2)$ to $O(nk)$. Yang et al. [32] introduced a Local Feature Statistics Histogram (LFSH). LFSH is obtained by computing the statistical data of three localized constant features including local depth, deviation angle between normals, and the projected point density. Frome et al. [14] proposed 3D Shape Context Signature (3DSC) that creates a spherical structure centered at a key point. The sphere is then split into a set of logarithmically spaced sub-spaces by dividing the sphere along the azimuth, elevation, and the radial directions, it then accumulates a weighted count of every point density that lies within a histogram. Note that 3DSC is not fully rotation invariant as it still has a degree of freedom in the azimuth direction. Blokland and Theoharis [33] introduced Radial Intersection Count Image (RICI) descriptor. Similar to spin-image, the RICI descriptor accumulates 3D points into a 2D histogram of integers. However, the difference is that spin-image accumulates projected point samples to create an estimate of the local surface area intersecting each bin, while RICI stores the counts of the intersections of circles with the surface points of the scene (each bin is an integer).

For LRF-based descriptors, they establish a local reference frame at a chosen key point for local surface feature encoding. Tombari et al. [34] proposed Unique Shape Context descriptor (USC), which is an extension of the 3DSC descriptor. USC provides an LRF to each 3DSC feature point, which improves the overall performance of the descriptor and is fully rotation invariant. Sukno et al. [35] presented rotational invariant 3D shape contexts using asymmetry patterns. It first computes the 3DSC descriptor then extracts asymmetry patterns from it. Similar to 3DSC, the Signature of Histogram of Orientation (SHOT) [10] also encodes local information about the surface in a spherical structure by splitting the sphere into sub-spaces. For every sub-space, a statistical histogram is computed. Malassiotis et al. [36] introduced SNAPSHOT, a local descriptor that captures images of the surface over each point

using a signed projected distance cue. Guo et al. [37] proposed Rotational Projection Statistics (RoPS) descriptor by first rotating the surface and then projecting the surface onto a 2D plane, the accumulated statistical information of those projected points are encoded into a 1D feature vector. Yang et al. [38] introduced the Triple Orthogonal Local Depth Image (TOLDI) descriptor, which is obtained from a sequence of local depth images captured with respect to the view plane of their proposed LRF into a feature vector.

3D local binary descriptors and learned descriptors are also worth mentioning here. Binary descriptors are composed by binary codes for accelerating matching and storage. B-SHOT descriptor [39] is a binary variant of the SHOT descriptor, which converts an 352-byte SHOT feature descriptor to an 352-bit descriptor. Blokland and Theoharis [40] proposed the Quick Intersection Count Change Image (QUICCI) descriptor. In contrast to their previous work [33] that stores integers representing intersection counts, the QUICCI descriptor stores booleans representing changes in intersection counts. QUICCI is resistant to clutter, as well as efficient to storage and matching. Rotational Contour Signatures (RCS) [13] addresses both real-valued and binary 3D local shape description problems, which is based on the contour information provided by a continually rotated object on a local reference frame (LRF) to encode the geometry.

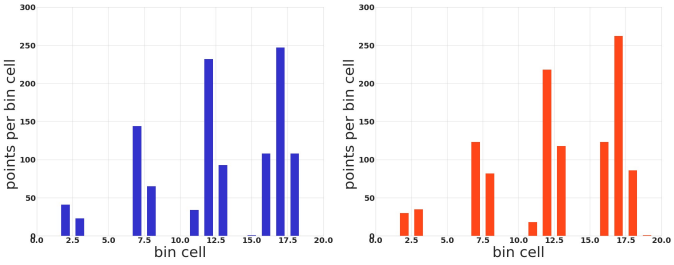
Learned descriptors use machine learning algorithms to learn a patch of local descriptors to match 3D data between two scenes. 3DMatch [41] uses a 3D convolutional neural network that learns to map a volumetric 3D patch around a key point on a 3D surface to an n-dimensional feature representation. Point Pair Feature Network (PPF-Net) [42] is a deep learning architecture that learns a local descriptor based on PPF geometric information and is highly aware of the global context. PPF-Net uses an N-to-N contrastive loss that correctly learns to solve combinatorial problems. SpinNet [43] is a cylindric convolutional neural network that extracts features from a spatially transformed local surface mapped into a cylindrical volume. Although learned descriptors present good performance on point cloud matching, they usually require a large amount of training data for efficient optimization. This may not be feasible in real-world applications because of the lack of enough training data. For this reason, we still focus on traditional descriptors.

Although the aforementioned descriptors have significantly contributed to the research community, they still struggle to achieve a balanced performance in the presence of noise, clutter, occlusion, and limited overlap. In this paper, we present a Dual-SI method that is simultaneously robust to a variety of nuisances.

3. Dual spin-image

In this section, we present our Dual-SI descriptor, a spin-image variant that encodes bi-directional information between a key point and its neighbors.

A good local feature descriptor should exhibit the following characteristics. The first is good descriptiveness. It is achieved by encoding the geometric information of the local surface in a bi-directional fashion between a key point and its neighbors within a support radius. As a result, we are able to capture richer geometric information (see Fig. 1 and Fig. 3) and achieve better matching results (see Sec. 4). The second is the robustness to occlusion and self-occlusion. We have designed a multi-scale support radius module to address the occlusion problem. The third is the robustness to noise and mesh resolution variation. We alleviate this issue by making use of a weighted kernel.



(a) S_1 number of points per bin (b) S_2 number of points per bin

Fig. 1. Illustration of the number of encoded points in a local area per bin cell obtained by interchanging the oriented point between a center point p and the its neighbor q_i. (a) represents the repartition of points over bin cells on the basis $(p, \vec{v_p})$, and (b) represents the repartition of points over bin cells on the basis $(q, \vec{v_q})$.

3.1. Bi-directional encoding

Given a point cloud with N points, we aim to encode a local surface of that point set within a radius r into a 2D histogram that is highly descriptive. We follow the spin-image procedure [16] to generate the 2D histogram described in Sec. 2.1. To capture more information about the geometry of a local surface shape, LRF based approach [10, 37] constructs a new coordinate system that generates three vectors $\{\vec{v_1}, \vec{v_2}, \vec{v_3}\}$ at the center point p, this may lead to a more comprehensive description of the local surface shape, while the stability of LRFs cannot be guaranteed. Unlike LRF that generates a single coordinate system with three axes, our approach consists of improving the relationship between a center point and all its neighbors within the range of a support radius length. Every local point set is defined by a center point p represented by its coordinates $p = \{x, y, z\} \in \mathbb{R}^3$, the radius r, and the center point neighbors $\{q_1, q_2, q_3, \cdots, q_N\}$, where $q_i = \{x_i, y_i, z_i\} \in \mathbb{R}^3$ denotes the i^{th} neighbor coordinates. By interchanging the position of the oriented point between the point p and its i^{th} neighbor, we can generate two signatures for every (p, q_i) pair using 2D coordinate of two different bases. Each basis corresponds to an oriented point/neighbor point p with its associated normal vector $\vec{v}$ perpendicular to the tangent plan P. The first basis is defined by the orientated point at p, where the 2D coordinates are $(p, \vec{v_p})$; the second basis is with the oriented point at q_i neighbor, where the 2D coordinates is $(q, \vec{v_q})$. The vectors $\vec{v_p}$ and $\vec{v_q}$ are normals of points p and q_i. The generated signatures are then defined as $S_1 = \left\{(p, \vec{v_p}); q_i\right\}$ and $S_2 = \left\{(q_i, \vec{v_q}); p\right\}$. This approach is motivated by the observation in Fig. 1, where the number of points per bin cell varies from one signature to the other. Given that the bins describe the encoded geometrical information of a local surface, combining signatures of different points per bin

ratio of the same local surface can increase the descriptiveness of the descriptor as will be verified in our evaluation. Thus, from the 2D coordinates of two bases, we use two spin-maps to independently compute two spin-images. i.e., each spin-map in Eq. 5 and Eq. 6 generates a 2D feature descriptor, which are denoted by S_1 and S_2, respectively.

$$(q, \vec{v_q}) : (\alpha_q, \beta_q) = \sqrt{\|q_i - p\|^2 - (\vec{v_p}.(q_i - p))^2},\ \vec{v_p}.(q_i - p) \quad (5)$$

$$(p, \vec{v_p}) : (\alpha_p, \beta_p) = \sqrt{\|p - q_i\|^2 - (\vec{v_q}.(p - q_i))^2},\ \vec{v_q}.(p - q_i) \quad (6)$$

The generated feature descriptors of size $w \times h$ are concatenated into a new 2D feature vector of size $2w \times h$ as shown in Fig. 2. Fig. 3 shows 1) a traditional spin-image S_1 where the oriented point is at key point p, 2) a traditional spin-image S_2 with the oriented point at q_i neighbor, and 3) the concatenation of them as the $f_{dual-si}$ feature.

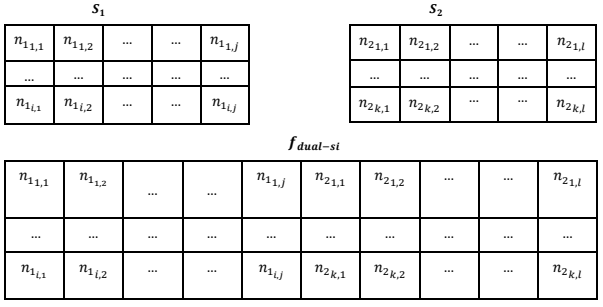


Fig. 2. Illustration of concatenating the generated two spin-images from two different bases.

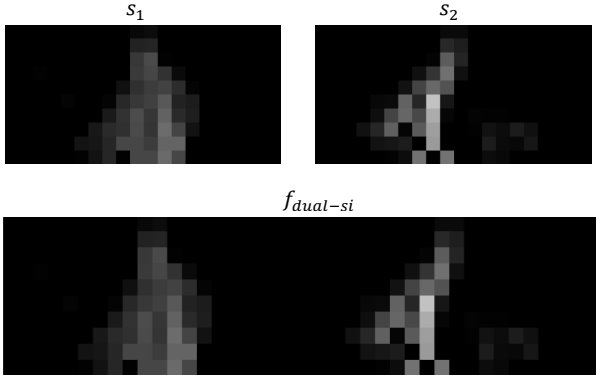


Fig. 3. Concatenating real generated spin-images from two different bases. Brighter pixels are bins with denser points.

Matching experiments on the B3R dataset have been conducted to observe how the generated spin-image signatures of size $w \times h$ behave when concatenated into a $2w \times h$ feature descriptor, as compared with the case of merging into a new $w \times h$ 2D feature descriptor. The illustration of the process of merging two signatures is shown in Fig. 4, where overlapping bins with point count being greater than zero are averaged. Fig. 5 shows the merging results with real spin-images. Note that we finally choose to concatenate two spin-images as will be verified in Sec. 4.2.

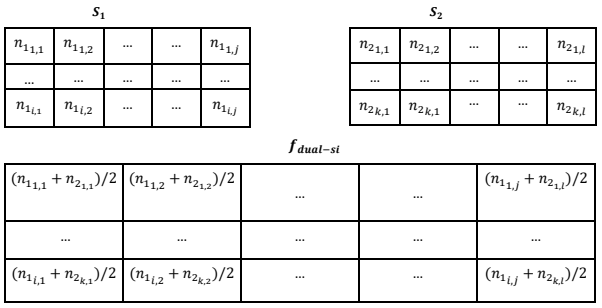


Fig. 4. Illustration of merging the generated two spin-images from two different bases.

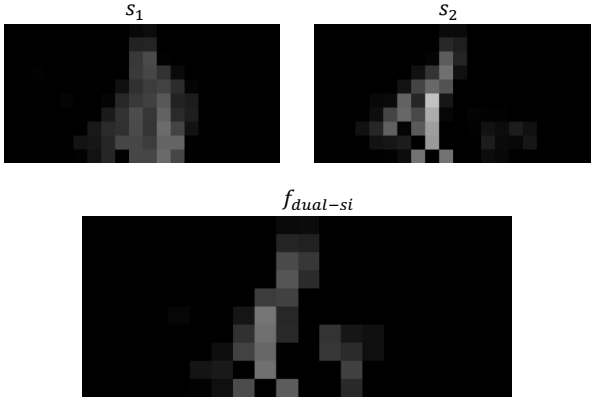


Fig. 5. Merging real generated spin-images from two different bases.

3.2. Multi-scale support radius

One of the characteristics of feature descriptors is locality. A feature occupies a relatively small area of the point cloud, thus, making the descriptor robust to clutter and occlusion. To implement this concept a local area is selected by defining a region within a radius r centered at a point p. To the best of our knowledge, no method has been proposed in the literature to determine the correct radius size to be used relative to ensure good performance for different scenarios.

A common strategy is to define a small region to steer clear of occlusion or self-occlusion hurdle. However, a small radius size may cause a descriptive paucity or object confusion due to partial similarities between two objects, therefore providing poor performance on object recognition and registration. A rather big radius size on the other hand may be subject to occlusion in a scene as shown in Fig. 6. To address this trade-off, our proposed Dual-SI method uses a multi-scale radius approach. Our approach differs [18] in three aspects: 1) the bi-directional encoding characteristic of our Dual-SI; 2) instead of choosing fewer points as we increase the radius length and the feature vector, we increasingly weight the generated Dual-SI as we decreased the support radius length; 3) we blend our dual spin-images with the weighted sum of each overlapping bins to have a final Dual-SI as shown in Fig. 7. Fig. 8 shows the Dual-SI generation procedure.

We empirically selected three support radius lengths r_1, r_2, r_3 with $(r_1 < r_2 < r_3)$, therefore, three Dual-SI feature descriptors f_1, f_2, f_3 (Fig. 7(a-c)) are generated. Each generated feature is

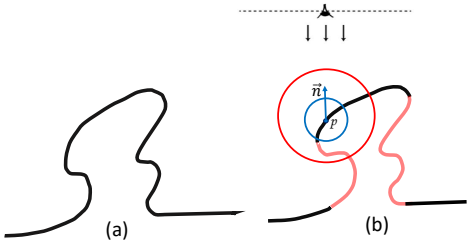


Fig. 6. Illustration of loss of information relative to the size of the support radius and occlusion areas (the small radius in blue does not contain any occluded region, while the radius in red contains occluded region). (a) A 2D shape. (b) The captured area (shown in dark) and the hidden area (shown in red).

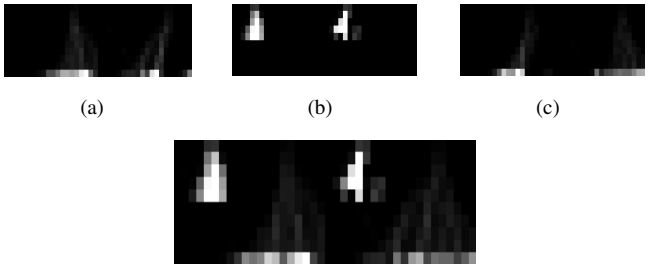


Fig. 7. Scaled Dual-SI. (a) shows Dual-SI with radius r_1, (b) shows Dual-SI with radius r_2, (c) show Dual-SI with radius r_3, (d) shows the merged 3 Dual-SIs generated with different scales.

weighted by a α_i parameter, with $\alpha_i \in \mathbb{R}$ such that, $F_i = \{\alpha_i f_i\}$. The final Dual-SI feature descriptor (Fig. 7(d)) is constructed by combining the three feature descriptors through:

$$F_{Dual-SI}(F_1, F_2, F_3) = \sum_{i=1}^{k} \alpha_i f_i \tag{7}$$

where $\sum \alpha_i = 1$ and $k = 3$ is the number of support radius. Throughout experiments we have found that better results were obtained when $\alpha_1 > \alpha_2 > \alpha_3$, i.e., the weight decrease as the radius length increase. The detailed multi-scaled radius algorithm for Dual-SI is presented in Algorithm 1.

3.3. Addressing noise problem

3D real data are usually noisy. To cope with the noise problem, we use a weighted kernel to remove sharp details and noise. The kernel used is a Gaussian filter as defined in Eq. 8, which operates a 2D convolution on the generated Dual-SI feature vector. The pixel values of the Dual-SI feature vector are normalized to [0,1] before convolution and denormalized after the filter is applied.

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{8}$$

For every bin cell b_{ij} in the Dual-SI, the corresponding applied weight is defined as:

$$W_{ij} = \frac{\left(2r - \sum_{i=1}^{n} \|p - q_i\|\right)^2}{N} \tag{9}$$

where n is the number of points recorded in a single bin cell of the feature vector and N the total number of non-zero normal points used to compute the Dual-SI. Fig. 8 (g)-(i) illustrates the filter used in our Dual-SI feature descriptor.

Algorithm 1 Dual-SI generation with multiple scales

Require: p: the center point; r_i: the support radius length; f_k: 2D feature descriptor representing Dual-SI; α_i: the weighted parameters applied to the bin cells $b_{k_{ij}}$ of the feature descriptor f_k; F: final Dual-SI feature descriptor

Ensure: The three Dual-SI feature descriptors f_1, f_2, f_3

Initialize F: as empty vector of size $w \times h$
for $i \leftarrow 0$ to w **do** ▷ w : width of F
  for $1 \leftarrow 0$ to h **do** ▷ h : height of F
    if $b_{2_{ij}} == 0$ and $b_{3_{ij}} == 0$ and $b_{1_{ij}} \neq 0$ **then**
      $F_{ij} \Leftarrow b_{1_{ij}}$ ▷ no weight applied
    else if $b_{3_{ij}} == 0$ and $b_{1_{ij}} == 0$ and $b_{2_{ij}} \neq 0$ **then**
      $F_{ij} \Leftarrow b_{2_{ij}}$ ▷ no weight applied
    else if $b_{1_{ij}} == 0$ and $b_{2_{ij}} == 0$ and $b_{3_{ij}} \neq 0$ **then**
      $f_{dual-si} \Leftarrow b_{3_{ij}}$ ▷ no weight applied
    else if $b_{3_{ij}} == 0$ and $b_{1_{ij}} \neq 0$ and $b_{2_{ij}} \neq 0$ **then**
      $F_{ij} \Leftarrow \alpha_1 b_{1_{ij}} + \alpha_2 b_{2_{ij}}$ ▷ weight sum is 1
    else if $b_{2_{ij}} == 0$ and $b_{1ij} \neq 0$ and $b_{3ij} \neq 0$ **then**
      $F_{ij} \Leftarrow \alpha_1 b_{1_{ij}} + \alpha_3 b_{3_{ij}}$ ▷ weight sum is 1
    else if $b_{1_{ij}} == 0$ and $b_{2_{ij}} \neq 0$ and $b_{3_{ij}} \neq 0$ **then**
      $F_{ij} \Leftarrow \alpha_2 b_{2_{ij}} + \alpha_3 b_{3_{ij}}$ ▷ weight sum is 1
    else if $b_{1_{ij}} \neq 0$ and $b_{2_{ij}} \neq 0$ and $b_{3_{ij}} \neq 0$ **then**
      $F_{ij} \Leftarrow \alpha_1 b_{1_{ij}} + \alpha_2 b_{2_{ij}} + \alpha_3 b_{3_{ij}}$ ▷ weight sum is 1
    end if
    $i = i + 1$
    $j = j + 1$
  end for
end for

4. Evaluation of the Dual-SI descriptor

In this section, we thoroughly evaluate the performance of our Dual-SI descriptor on standard datasets including the Bologna 3D Retrieval (B3R) [19], the Bologna Dataset5 (BoD5) [10], the Bologna Mesh Registration (BMR) [10], the UWA 3D Object Recognition (U3OR) [20, 21], and the UWA 3D Modeling (UWA3M) [22] datasets. Additionally, a thorough comparison of our Dual-SI results to several state-of-the-art methods is conducted to demonstrate the overall superiority of the Dual-SI descriptor. Finally, we show the proficiency of our Dual-SI descriptor in point cloud registration application on Microsoft Kinect and Konica Minolta data. The Dual-SI is implemented in C++ language. The state-of-the-art methods used for comparison against our Dual-SI are taken from the point cloud library (PCL) [44]. The experiments are conducted on an Intel Xeon E3-1220 desktop with a 3.5GHZ CPU and 8GB of

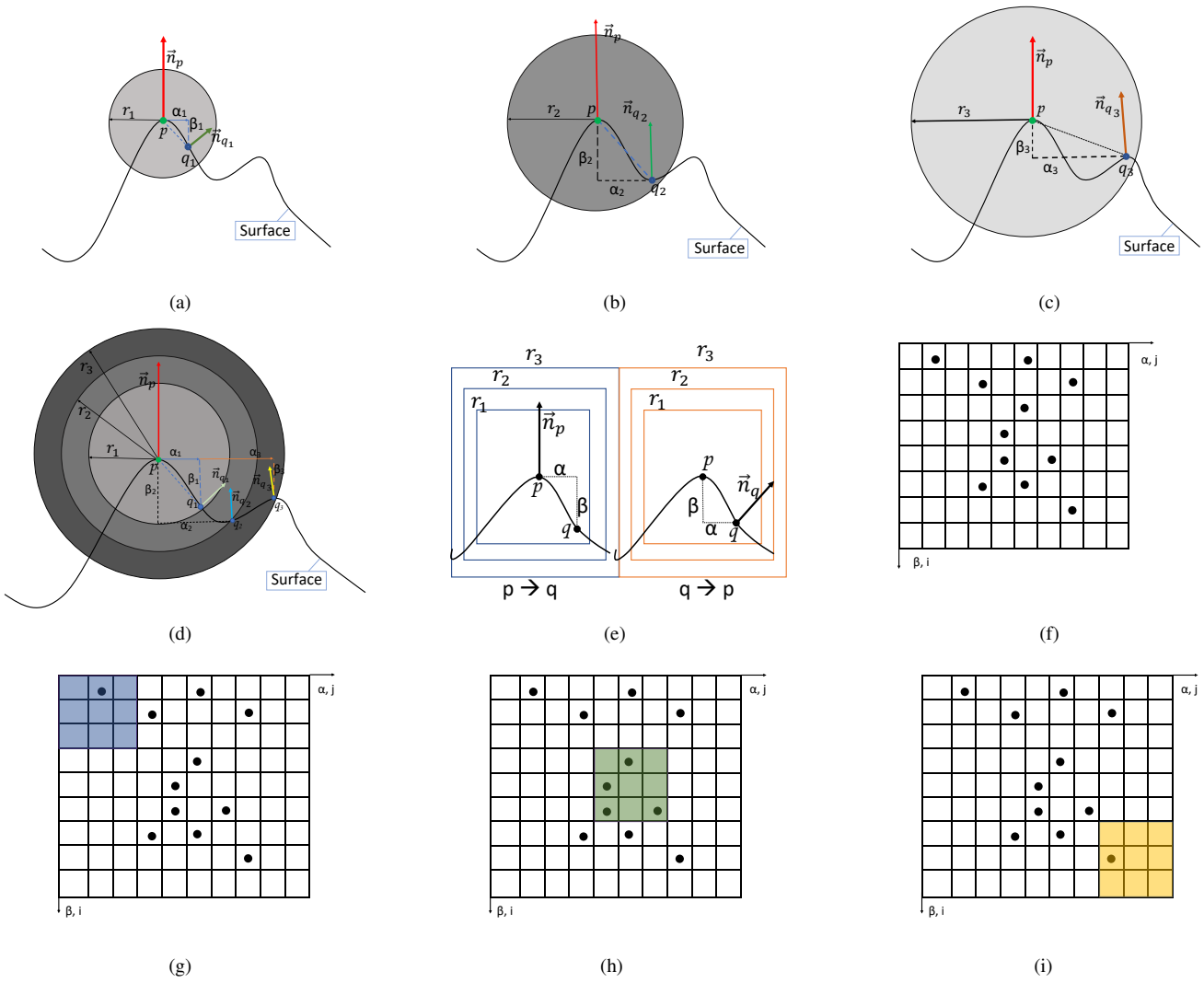


Fig. 8. Illustration of Dual-SI generation process. (a)-(d) generate the dual spin-image with a multi-scale radius. (e) concatenates both spin-images (with normal at p center point and q_i neighbor point), then blended by α (blending weight for each radius) for the three generated Dual-SIs. (f) illustrates binning process of the 3D points to a 2D space (the spin-image). (g)-(i) apply a weighted kernel to the Dual-SI to achieve robustness to noise.

RAM. No parallel computing or GPU processing was used for the Dual-SI implementation.

4.1. Experimental setup

Prior to the evaluation, we discuss the technical aspects and details about the implementation of experiments, including the dataset description, parameter settings, and evaluation criteria.

4.1.1. Datasets

The selected datasets to conduct experiments are public datasets with different application scenarios, i.e., B3R dataset for shape retrieval, the UWAOR and BoD5 datasets for 3D object recognition, the BMR and the UWA3M datasets for partial view matching. Samples of the described datasets are shown in Fig. 9. The number of matching pairs in each dataset is listed in Table 2. The ground-truth transformations between each model and its scenes instances were provided as a prior for all the aforementioned datasets.

B3R: the Bologna 3D Retrieval[1] is a generated synthetic dataset taken from the Stanford 3D Scanning Repository[2] [45] and designed for shape retrieval context. The B3R dataset contains 18 models and 6 scenes obtained from adding nuisance and rotating the models. We have also created another group of scene objects with respect to Gaussian noise with a standard deviation of 0.1 mr independently added to the x, y, and z axis of the scene points, in order to test the performance of the feature descriptors in the presence of noise and analyze the effect of our kernel module on the Dual-SI.

BoD5 : the Bologna Dataset5 is a real-world dataset acquired with the Microsoft Kinect sensor. It contains 26 models scanned from 6 objects and 16 scenes (15 test scenes + 1 tuning scene). The models were taken from the Stanford 3D Scanning Repository [45].

[1] vision.deis.unibo.it/key points3d
[2] www.graphics.stanford.edu/data/3Dscanrep

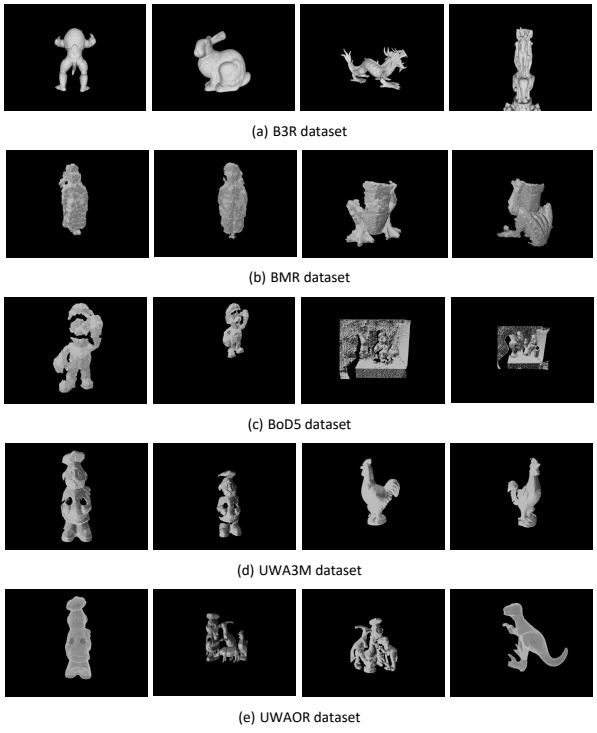


Fig. 9. Dataset samples taken from B3R, BMR, BoD5, UWA3M, and UWAOR.

BMR : the Bologna Mesh Registration is a real-world dataset acquired with the Microsoft Kinect sensor for some objects and the Microsoft Space Time technique for other objects. The models and scenes on this dataset are composed of several views of 6 objects (*Mario*, *Squirrel*, *Frog*, *Duck*, *Doll*, *PeterRabbit*).

UWA3M : UWA 3D Modeling dataset[3] contains 22, 16, 16, and 21 2.5D scans from four objects (*Chef*, *Chicken*, *Parasaurolophus*, and *T-Rex*), respectively. They address range image registration (view matching) scenarios. The 2.5D models were scanned from different views using the Konica Minolta Vivid 910 scanner from a single viewpoint, leading to feature description and matching in this dataset to be confronted with missing regions, holes, and self-occlusion problems. The ground-truth transformations for this dataset are obtained with manual alignment and iterative closest points (ICP) algorithms [46]. We created another three groups of the scenes data that respectively contain Gaussian noise with 0.1 mr, 0.3 mr and, 0.5 mr standard deviations to evaluate the stability of our descriptor on self-occluded and missing regions objects that contains noise as well.

UWAOR : the UWA 3D Object Recognition[4] is one of the most frequently used real-world dataset. UWAOR contains reconstructed models from several point clouds of objects acquired with the Minolta Vivid 910 scanner. It comprises 5 models (including, C*chef*, *Chicken*, *Parasaurolophus*, *T-Rex*, and *Rhino*) and 50 scenes. Each scene contains four to five models in the presence of clutter and occlusion. Similar to the UWA3M dataset, we derived three groups of scenes data from the UWAOR dataset that contains Gaussian noise with respectively 0.1 mr, 0.3 mr and, 0.5 mr standard deviations.

[3] stafhome.ecm.uwa.edu.au/ 00053650/3Dmodeling.html
[4] stafhome.ecm.uwa.edu.au/ 00053650/recognition.html

4.1.2. Evaluation criteria

The performance of our proposed Dual-SI are measured using the Recall vs 1-Precision curve (RPC) and the Area under the recall vs 1-precision Curve (AUC) metrics. AUC can measure the performance of a descriptor in an overall manner. RPC metric is widely used in 2D and 3D evaluation performance [47]. The RPC is calculated as follows. Given a model data, a scene data, and its corresponding ground truth transformation, a model feature is matched against all the scene features to find the closest and the second closest features. If the ratio between the smallest and second smallest distance is lesser than a chosen threshold, we consider the model and the corresponding scene feature to be a match. The match is further validated if the distance between the selected points is small enough, i.e., smaller than the support radius selected in our paper, otherwise, it would be considered as a false match. By changing the threshold values, a curve is then generated. The recall and 1-Precision are given by the following definition:

$$\text{recall} = \frac{\text{the number of correct matches}}{\text{total number of corresponding features}} \tag{10}$$

$$1 - \text{precision} = \frac{\text{the number of false matches}}{\text{total number of matches}} \tag{11}$$

In our experiment, we randomly selected 1000 points from every model as key points, and their corresponding points on the scene data are located using the ground truth transformation as mentioned in [37, 28]. Those evaluation criteria are computed between two surfaces to be matched, i.e., between a model and a scene, and are considered as valid experimental data only when the two surfaces have overlapping regions. In the selected datasets only the B3R and the UWAOR satisfy this norm due to their retrieval aspect and their model-based object recognition context. This is different for 3D registration. In the UWA3M dataset, not every two views in an object overlap with each other. In this case only the view pairs with at least 10% overlapping ratio (i.e., the ratio between the number of corresponding points and the minimum number of point counts of the two view pairs [22]) are considered.

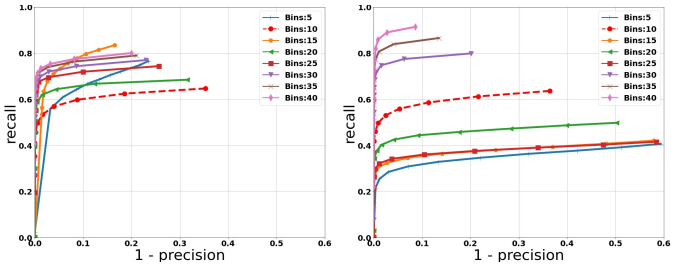
Given the generated recall vs 1-precision curve, we provide the AUC value of curves associated with each descriptor to measure the performance of a descriptor in an overall manner.

4.1.3. Compared methods

An important aspect of our experiments is the comparison of our Dual-SI with the state-of-the-art 3D local shape descriptors, including spin-image [16, 17, 24], Fast Point Feature Histogram (FPFH) [31], Signature Histogram of Orientation (SHOT) [10], SNAPSHOT [36], Rotational Projection Statistics (RoPS) [37], and the recent Rotational Contour Signature (RCS) [13] local descriptor. The parameter settings of the evaluated descriptors are presented in Table 1.

Table 1. Parameters settings and storage requirements of the tested descriptors. Note that the FPFH requires a smaller radius because its effective influence radius is 2r, and mr represents mesh resolution.

Descriptor	Support Radius (mr)	Dimensionality	Length	Storage (bit)
state-of-the-art descriptors and Dual-SI				
Spin-image	15	15×15	225	225×8
SNAPSHOT	15	40×40	1600	1600×8
FPFH	10	3×11	33	33×8
SHOT	15	8×2×2×11	352	352×8
RoPS	15	3×3×3×5	135	135×8
RCS	15	6×12	72	72×8
Dual-SI	12, 15, 17	15×15×2	450	450×8



(a) Concatenating feature vectors (b) Merging feature vectors

Fig. 10. Comparison between merging and concatenating the generated feature vector with different bin size.

4.2. Parameter analysis of Dual-SI

As previously stated in this work, Dual-SI generates two distinct spin-image signatures for each (p, q_i) pair, which are then combined to form the Dual-SI feature. We investigated various combinations of the resulting two spin-images with varying bin sizes, first by concatenating (Fig. 10 (a)) and then by merging (Fig. 10 (b)) both signatures. The B3R dataset is used for this experiment. The results of the experiments show that our method is less vulnerable to varying the bin size when the generated signatures are concatenated than merged. Merging both signatures offers the best results with larger bin size (30, 35, 40) while the concatenating operation achieves the best performance with a bin size of 15. To also ensure the compactness of Dual-SI (hopefully with less bins), we set the bin size to 15 and employ the concatenation operation for aggregating two spin-images for a point pair (p, q_i).

4.3. Descriptor matching performance

In this section, we report the results of feature matching experiments of Dual-SI and state-of-the-art algorithms conducted on the previously described datasets. Each dataset addresses different application scenarios including shape retrieval, 3D object recognition, and, partial view matching. The B3R dataset mainly studies the descriptiveness of feature descriptors. The UWAOR and BoD5 datasets evaluate the descriptors' robustness to clutter, occlusion and noise. Finally, the UWA3M and BMR datasets assess the strength of the descriptor on self-occluded objects, missing regions, holes and noise. In addition, we test the registration performance of feature descriptors under different root mean square error (RMSE) thresholds on the BMR, BoD5, UWAOR, and the UWA3M datasets using the Harris 3D (H3D) algorithm for key point selection.

Table 2. The number of matching pairs in datasets

	B3R	BMR	BoD5	UWA3M	UWAOR
Number of model objects	12	6	6	4	5
Number of scene objects	18	97	16	75	50
Number of matching pairs	18	485	43	496	188

4.3.1. Retrieval context: B3R dataset

The results of evaluated descriptors in terms of RPC and AUC (the numbers in the legend) on the B3R dataset are shown in Fig. 11(a-b) and (g-h).

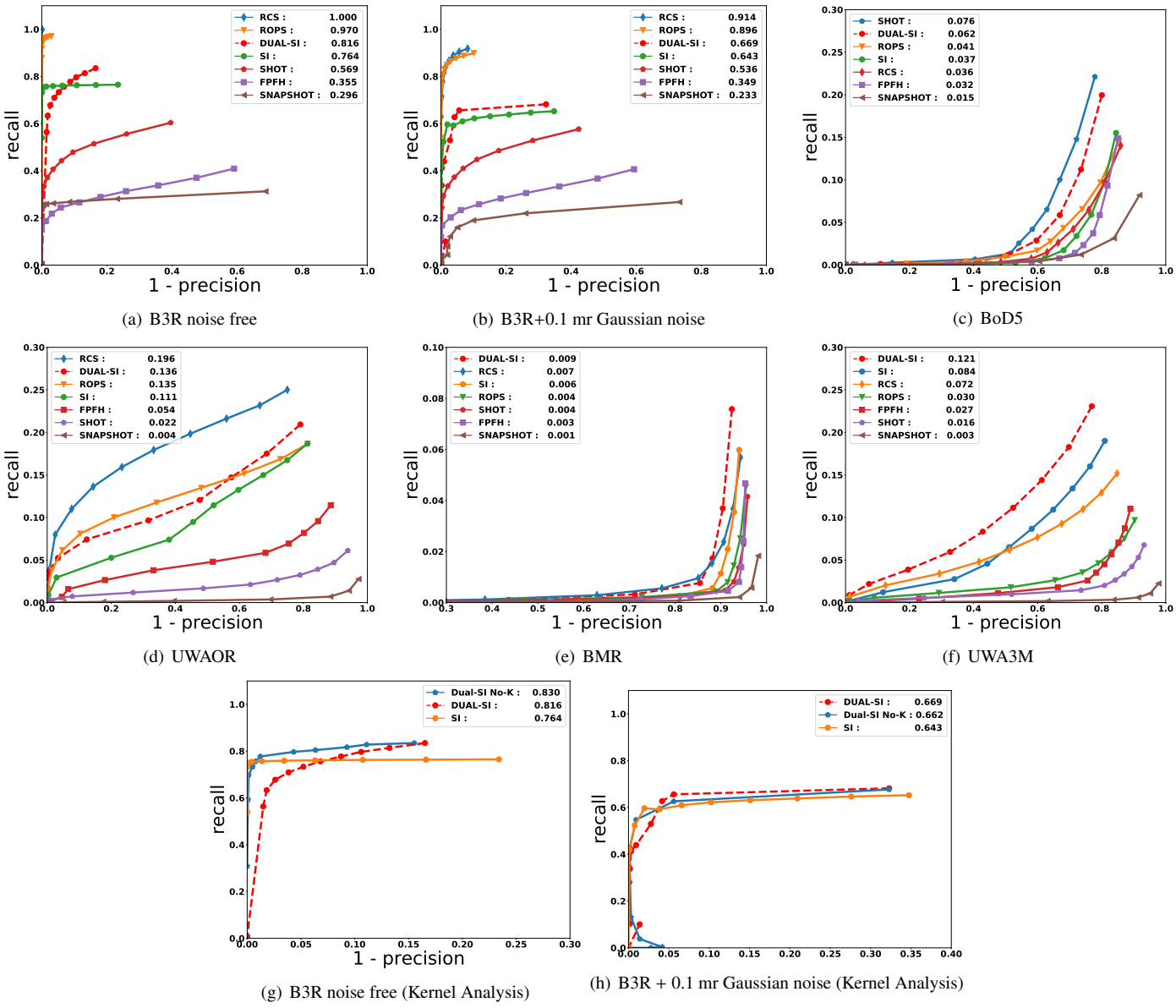
Two conclusions can be drawn from the figures. First, in the noise free B3R dataset, Dual-SI achieves the third best performance, surpassed by RCS and RoPS. It is clear that Dual-SI outperforms the traditional spin-image. Second, under the influence of noise (0.1mr Gaussian noise), the performance of all descriptors deteriorates. Given that our Dual-SI concatenates two distinct spin-image signatures, the impact of noise on our feature descriptor is not as obvious as that on the conventional spin-image. On that note, the Dual-SI outperforms the traditional spin-image on both recall and 1-precision metrics, validating the efficacy of our weighted kernel to smooth the spin-images in the presence of noise.

Fig. 11 (g-h) show the results of Dual-SI with and without our weighted kernel against spin-image. On the B3R noise free dataset, the Dual-SI with no kernel offers better performance than Dual-SI with the kernel module. Note that both Dual-SI versions outperform the traditional spin-image. On the B3R + 0.1 mr Gaussian noise, we observe that applying our kernel module improves the Dual-SI performance from 0.662 AUC value to 0.669 as compared with the Dual-SI variant with no kernel module. This can validate the effectiveness of the kernel module in the presence of noise.

4.3.2. 3D object recognition context: UWAOR and BoD5 datasets

The results of our experiments on the UWAOR are presented in Fig. 11(d) and Fig. 12. The figure clearly shows that the ranking of descriptors varies dramatically when compared to the results on the B3R dataset. Specifically, the SHOT descriptor that obtained acceptable performance on the B3R dataset drops to second worst one on the UWAOR dataset. By contrast, our Dual-SI descriptor achieves the second best result on this dataset. The same observation can be drawn on the UWAOR dataset with different levels of noise as shown in Fig. 12, which demonstrates that SI can simultaneously resist the impact of clutter, occlusion, and noise.

On the BoD5 dataset with real noise, the same observations can be made about the shift in performance order. The results of our experiments on the BoD5 are presented in Fig. 11(c). The SHOT descriptor offers the best performance against all the other descriptors followed by our Dual-SI, spin-image, FPFH, RoPS, RCS, and SNAPSHOT. The RCS descriptor that presents the best performance on the B3R and UWAOR datasets now produces the second worst performance on the BoD5 dataset.



(a) B3R noise free (b) B3R+0.1 mr Gaussian noise (c) BoD5

(d) UWAOR (e) BMR (f) UWA3M

(g) B3R noise free (Kernel Analysis) (h) B3R + 0.1 mr Gaussian noise (Kernel Analysis)

Fig. 11. Feature matching performance on the experimental datasets.

Above observations expose the difficulty to maintain stability when varying datasets and application contexts for 3D local descriptors. However, it can be noticed that our Dual-SI preserves top-ranked performance on B3R, UWAOR and BoD5 datasets.

4.3.3. 2.5D view matching context: UWA3M, BMR datasets

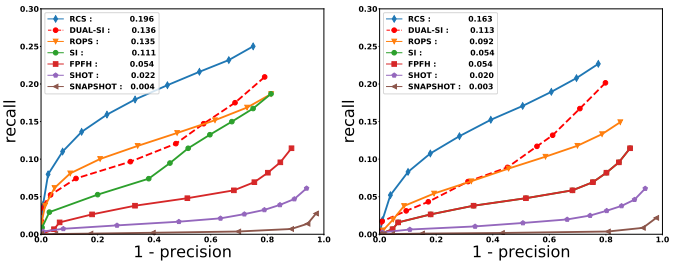
In addition to object recognition and shape retrieval context, we also conduct experiments on partial view matching datasets including UWA3M and BMR. 2.5D view matching has an important role for application such as 3D modeling [48] and object pose estimation [49]. The results are shown in Fig. 11(e-f) and Fig. 13.

On the UWA3M dataset, We can observe that our Dual-SI surpasses all competitors. Notably, the SHOT descriptor that offers the best performance on the BoD5 dataset, now presents the second worst result on the UWA3M dataset. When additionally adding noise to this dataset, Dual-SI still maintains the best performance. On the BMR dataset with real noise, as shown in Fig. 11(e), our Dual-SI still outperforms all compared descriptors by a large margin.

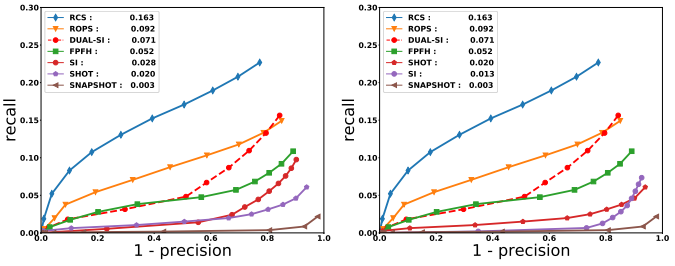
4.4. Point cloud registration performance

In addition to the feature matching experiments, we further test the 3D point cloud registration performance of the descriptors. The registration is conducted using the RANSAC [50] pipeline with respect to RMSE thresholds, where the initial correspondences are generated by matching local feature descriptors. The results are show in Fig. 14. The following observations can be made.

1) On the UWA3M dataset (Fig. 14(a)), the main findings are summarized as follows. First, Dual-SI and RoPS are the two top-ranked descriptors. Second, Dual-SI outperforms all the compared descriptors when the RMSE threshold is in the range of [0, 9] mr. Third, the RoPS descriptor achieves the best

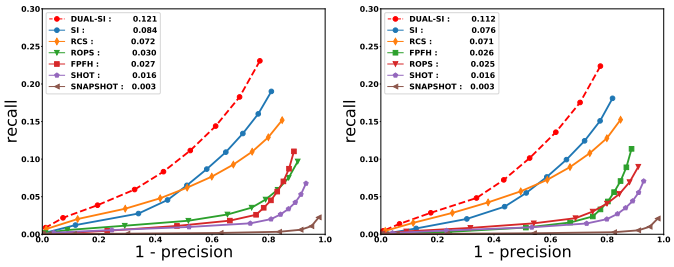


(a) UWAOR noise free (b) UWAOR+0.1 mr Gaussian noise

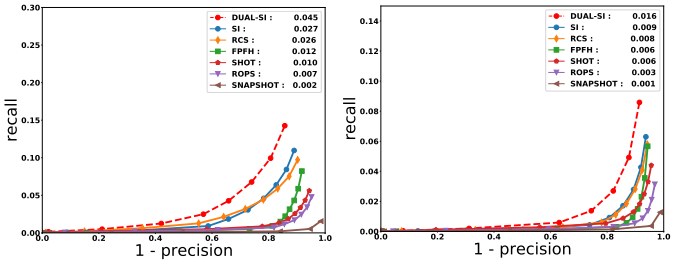


(c) UWAOR+0.3 mr Gaussian noise (d) UWAOR+0.5 mr Gaussian noise

Fig. 12. Feature matching performance on the UWAOR dataset with Gaussian noise.



(a) UWA3M noise free (b) UWA3M+0.1 mr Gaussian noise



(c) UWA3M+0.3 mr Gaussian noise (d) UWA3M+0.5 mr Gaussian noise

Fig. 13. Feature matching performance on the UWA3M dataset with Gaussian noise.

performance as the RMSE threshold further increases, and our Dual-SI takes the second place.

2) On the BMR dataset (Fig. 14(b)), we are able to observe the change in performance when cross dataset that also occurred in feature matching experiments. Three salient observations can be made from the figure. First, the SHOT descriptor offers the best registration accuracy result and outperforms all the descriptors by a large margin, followed by Our Dual-SI, FPFH, spin-image, RoPS and SNAPSHOT. Second, four of the six descriptors used in this experiment (including Dual-SI, FPFH, RoPS and spin-image) yield similar results. Finally, the SNAPSHOT presents very poor performance, whose registration accuracy is smaller than 10%.

3) On the UWAOR dataset (Fig. 14(c)), one can see that our Dual-SI presents the best performance, and outperforms all the descriptors by a large margin, followed by RoPS, SHOT, FPFH, spin-image, and SNAPSHOT. This further verifies the effectiveness of Dual-SI on coping with the occlusion problem.

4) On the BoD5 dataset (Fig. 14(d)), our Dual-SI achieves the third best performance. Overall, we find that the performance of Dual-SI remains relatively stable when aligning point clouds in different scenarios.

The average time costs of tested descriptors for 3D registration on four datasets is given in Fig. 15. Overall, the time cost of all tested descriptors achieve comparable performance.

4.5. Computational efficiency

Besides feature matching descriptor performance, we have also tested the computational efficiency of our Dual-SI against state-of-the-art feature descriptors with different support radii. The number of points in the local surface area is proportional to the support radius length. The purpose of this experiment is to cover application cases that may require different scales of descriptors. The results (all the methods are implemented with C++) are shown in Fig. 16.

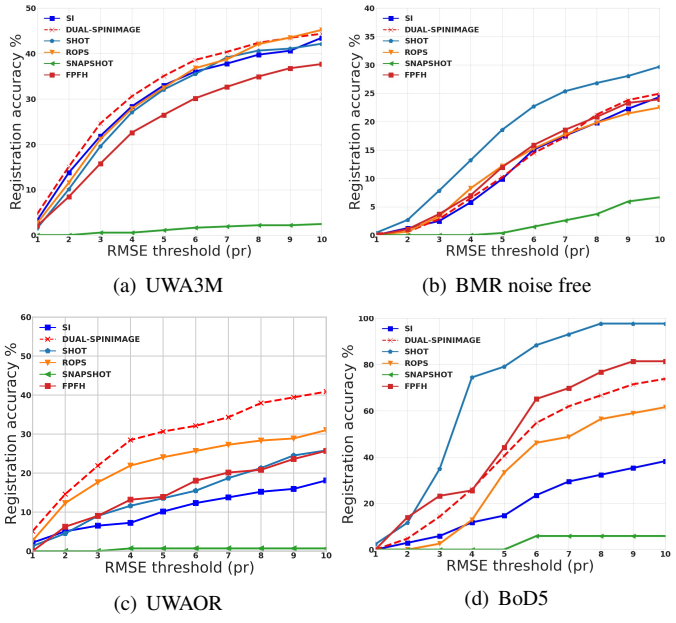
Several observations are notable and summarized as follows. First, the computational time cost increases proportionally to the the support radius length due to the increase of point counts. Second, the RCS descriptor [13] is shown to be the most efficient one, followed by the original spin-image, SHOT, SNAPSHOT, our Dual-SI, FPFH, and RoPS descriptors. The computation time of RoPS increases drastically when the support radius passes 15 mr. Regardless of the multi-scale aspect of our Dual-SI, its computational time efficiency still outperforms FPFH and RoPS descriptors. For every radius r_1 in the scale, our Dual-SI uses two more radii $r_0 = r_1 - 3$ mr and $r_2 = r_1 + 2$ mr.

4.6. Applications

Finally, we present some point clouds matching examples with our Dual-SI feature descriptor. We consider the UWA3M and the U3OR datasets acquired with the Konica Minolta Vivid 910 scanner and the Microsoft Kinect sensor for this experiment, respectively. For the UWA3M dataset, we consider the *buddha* and the *chicken* view pairs. For the U3OR dataset, we consider scenes that match the *buddha*, the *chicken* and the *t-rex* view pairs to their corresponding models. RANSAC is employed to find consistent matches. It is worth noting that point clouds obtained with the Kinect sensor are sparse and noisy. The matching results are shown in Fig. 17, which suggest that Dual-SI is able to produce sufficient consistent feature matches between point clouds with limited overlap, and is robust to data modality changes.

4.7. Limitations of Dual-SI

Above experiments indicate that Dual-SI achieves top-ranked performance almost on all datasets, in both feature matching



(a) UWA3M (b) BMR noise free

(c) UWAOR (d) BoD5

Fig. 14. 3D rigid registration accuracy performance using the Dual-SI descriptor with RANSAC estimator on four experimental datasets with respect to different RMSE thresholds.

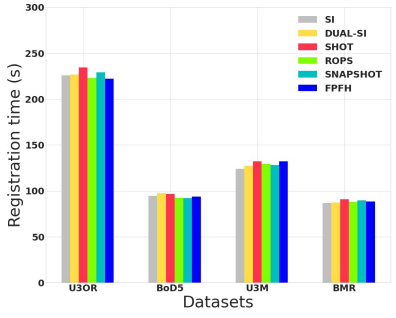


Fig. 15. Average time cost for registering a shape pair on four datasets.

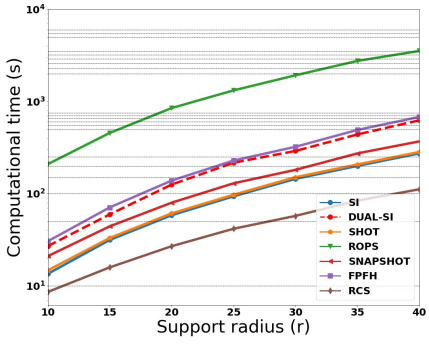


Fig. 16. Time costs of computing matches with recall vs 1-precision metric of the Dual-SI feature descriptor and state-of-the-art descriptors with respect to different support radii.

and 3D point cloud registration experiments. This suggests that Dual-SI is a descriptor with balanced performance in the presence of data modality change, noise, occlusion, clutter, holes, limited overlap, and missing regions. Even though, Dual-SI still has the following limitations that may be further improved in future works.

1) Dual-SI is grounded on oriented point pairs. In other words, normals are usually required for Dual-SI calculation. Similar to many normal-based descriptors such as spin-image and FPFH, the stability of Dual-SI is depended on the quality of normals. We can find that on Kinect datasets such as BMR and BoD5, the performance of Dual-SI also clearly degrades due to noisy normals.

2) Dual-SI is suitable for 3D rigid data and may encounter difficulties when used for deformable objects. In future studies, intrinsic geodesics can be used with the Dual-SI feature descriptor to obtain isometric invariance. Similar to spin-image, Dual-SI is optimal for objects with rich shape and geometric information. Symmetrical or planar objects, such as balloons and walls, are not covered by these types of features. This issue of Dual-SI could be further addressed by combining shape and color information [34].

3) The compactness performance of Dual-SI is not outstanding (450 dimensional). This may lead to less efficient matching and storage. We expect to further shortening the length of Dual-SI with unsupervised learning such as PCA and auto-encoder.

5. Conclusions and future work

In this paper, we presented a novel Dual-SI feature for 3D local surface description. The key components of our proposed method include 1) a bi-directional approach between a key point and its neighbors to boost the feature's distinctiveness, 2) a multi-scale strategy to achieve robustness to occlusion and clutter, and 3) a weighted kernel to smooth the 2D feature representation of the local surface and reduce the impact of noise. We thoroughly conducted experiments on five challenging datasets that contain a variety of nuisances including noise, occlusion, clutter, holes, limited overlap, and missing regions. The conducted feature matching experiments in Sec. 4.3 and 3D registration experiments in Sec. 4.4 show that 1) our Dual-SI feature descriptor behaves very stable cross different datasets and 2) achieves top-3 performance on all datasets.

Regarding the future work on Dual-SI, we consider investigating bi-directional local reference frame (LRF), i.e., constructing a local reference frame at a key point and for every neighbor. This approach has the potential to improve the descriptiveness and robustness of the local feature descriptor by providing us with six axes to encode its geometry information. On public datasets, a quantitative study of local shape description revealed that LRF-based approaches are more efficient than LRF-independent approaches [47].

References

[1] 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 2014;36:2270–2287.

[2] Osada, R, Funkhouser, T, Chazelle, B, Dobkin, D. Shape distributions. ACM Transactions on Graphics 2002;21:807–832.

[3] Gao, Y, Dai, Q, Zhang, NY. 3D model comparison using spatial structure circular descriptor. Pattern Recognition 2010;43:1142–1151.

[4] Rusu, RB, Bradski, G, Thibaux, R, Hsu, J. Fast 3D recognition and pose using the viewpoint feature histogram. In: Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems. 2010, p. 2155–2162.

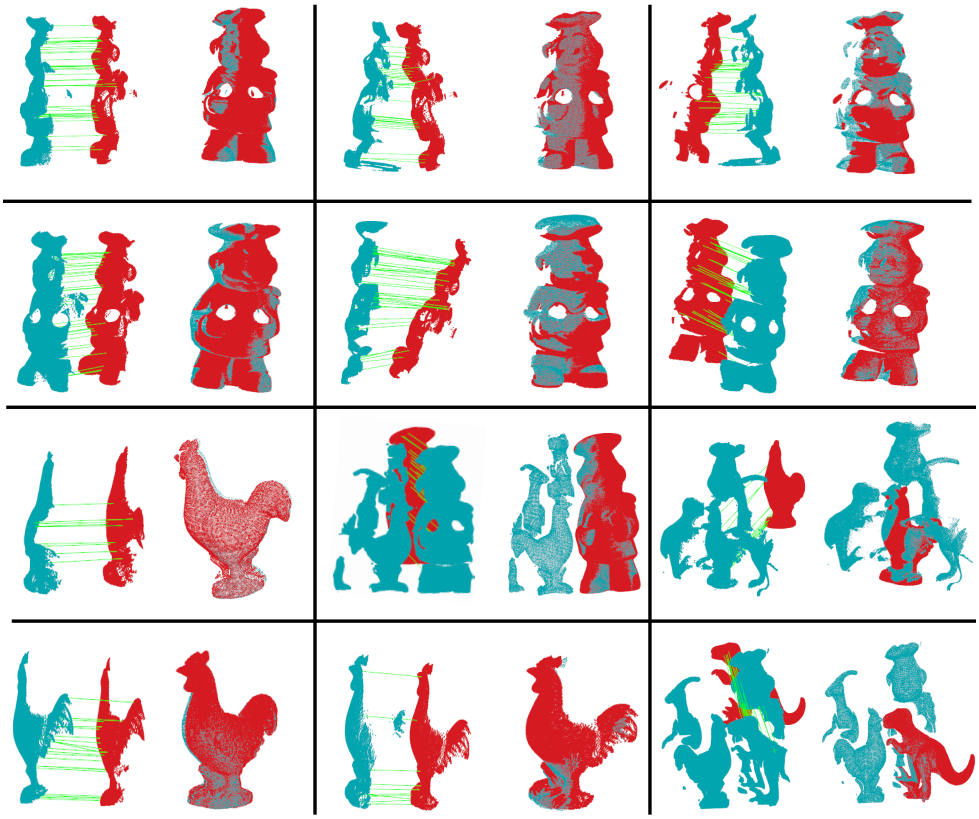


Fig. 17. Feature matching and registration examples on Kinect and the Konica Minolta Vivid point clouds based on our proposed Dual-SI.

[5] Aldoma, A, Vincze, M, Blodow, N, Gossow, D, Gedikli, S, Rusu, RB, et al. Cad-model recognition and 6dof pose estimation using 3D cues. In: IEEE International Conference on Computer Vision Workshops (ICCV Workshops). 2011, p. 585–592.

[6] Wohlkinger, W, Vincze, M. Ensemble of shape functions for 3D object classification. In: Proc. IEEE International Conference on Robotics and Biomimetics. 2011, p. 2987–2992.

[7] Shah, SAA, Bennamoun, M, Boussaid, F. A novel 3D vorticity based approach for automatic registration of low resolution range images. Pattern Recognition 2015;48:2859–2871.

[8] Ovsjanikov, M, Mérigot, Q, Mémoli, F, Guibas, L. One point isometric matching with the heat kernel. Computer Graphics Forum 2010;29:1555–1564.

[9] Lei, Y, Guo, Y, Hayat, M, Bennamoun, M, Zhou, X. A two-phase weighted collaborative representation for 3D partial face recognition with single sample. Pattern Recognition 2016;52:218–237.

[10] Tombari, F, Salti, S, Di Stefano, L. Unique signatures of histograms for local surface description. In: Proc. European Conference on Computer Vision; vol. 6313. 2010, p. 356–369.

[11] Chua, CS, Jarvis, R. Point signatures: A new representation for 3D object recognition. International Journal of Computer Vision 1997;25:63–85.

[12] Bariya, P, Novatnack, J, Schwartz, G, Nishino, K. 3D geometric scale variability in range images: Features and descriptors. International Journal of Computer Vision 2012;99:232–255.

[13] Yang, J, Zhang, Q, Xian, K, Xiao, Y, Cao, Z. Rotational contour signatures for both real-valued and binary feature representations of 3D local shape. Computer Vision and Image Understanding 2017;160:133–147.

[14] Frome, A, Huber, D, Kolluri, R, Bülow, T, Malik, J. Recognizing objects in range data using regional point descriptors. In: Proc. European Conference on Computer Vision; vol. 3023. 2004, p. 224–237.

[15] Rusu, RB, Marton, ZC, Blodow, N, Beetz, M. Persistent point feature histograms for 3D point clouds. In: Proc. International Conference on Intelligent Autonomous Systems. 2008, p. 119–128.

[16] Johnson, AE, Hebert, M. Surface matching for object recognition in complex three-dimensional scenes. Image and Vision Computing 1998;16:635–651.

[17] Johnson, AE, Hebert, M. Using spin images for efficient object recognition in cluttered 3D scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence 1999;21:433–449.

[18] Dinh, HQ, Kropac, S. Multi-resolution spin-images. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition; vol. 1. 2006, p. 863–870.

[19] Tombari, F, Salti, S, di Stefano, L. Performance evaluation of 3D keypoint detectors. International Journal of Computer Vision 2013;102:198–220.

[20] 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 2010;89:348–361.

[21] Mian, AS, Bennamoun, M, Owens, R. Three-dimensional model-based object recognition and segmentation in cluttered scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence 2006;28:1584–1601.

[22] Mian, AS, Bennamoun, M, Owens, RA. A novel representation and feature matching algorithm for automatic pairwise registration of range images. International Journal of Computer Vision 2006;66:19–40.

[23] Johnson, AE, Hebert, M. Efficient multiple model recognition in cluttered 3-d scenes. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1998, p. 671–677.

[24] Johnson, AE. Surface landmark selection and matching in natural terrain. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition; vol. 2. 2000, p. 413–420.

[25] Ruiz Correa, S, Shapiro, LG, Meliă, M. A new signature-based method for efficient 3-d object recognition. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition; vol. 1. 2001, p. I–769.

[26] Darom, T, Keller, Y. Scale-invariant features for 3-d mesh models. IEEE Transactions on Image Processing 2012;21:2758–2769.

[27] Lu, R, Zhu, F, Hao, Y, Wu, Q. Simple and efficient improvement of spin image for three-dimensional object recognition. Optical Engineering 2016;55:113102.

[28] 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 2015;293:196–213.
[29] Liang, L, Wei, M, Szymczak, A, Pang, WM, Wang, M. Spin contour. IEEE Transactions on Multimedia 2016;18:2282–2292.
[30] Kokkinos, I, Bronstein, MM, Litman, R, Bronstein, AM. Intrinsic shape context descriptors for deformable shapes. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. 2012, p. 159–166.
[31] Rusu, RB, Blodow, N, Beetz, M. Fast point feature histograms (fpfh) for 3D registration. In: Proc. IEEE International Conference on Robotics and Automation. 2009, p. 3212–3217.
[32] Yang, J, Cao, Z, Zhang, Q. A fast and robust local descriptor for 3D point cloud registration. Information Sciences 2016;346:163–179.
[33] van Blokland, BI, Theoharis, T. Radial intersection count image: A clutter resistant 3d shape descriptor. Computers & Graphics 2020;91:118–128.
[34] Tombari, F, Salti, S, Di Stefano, L. Unique shape context for 3D data description. In: Proc. ACM Workshop on 3D Object Retrieval. 2010, p. 57–62.
[35] Sukno, FM, Waddington, JL, Whelan, PF. Rotationally invariant 3D shape contexts using asymmetry patterns. Proc International Conference on Computer Graphics Theory and Applications 2013 2013;:7–17.
[36] Malassiotis, S, Strintzis, MG. Snapshots: A novel local surface descriptor and matching algorithm for robust 3D surface alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence 2007;29:1285–1290.
[37] 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 2013;105:63–86.
[38] Yang, J, Zhang, Q, Xiao, Y, Cao, Z. Toldi: An effective and robust approach for 3D local shape description. Pattern Recognition 2017;65:175–187.
[39] Prakhya, SM, Liu, B, Lin, W. B-shot: A binary feature descriptor for fast and efficient keypoint matching on 3D point clouds. In: Proc. IEEE International Conference on Intelligent Robots and Systems; vol. 2015-December. 2015, p. 1929–1934.
[40] van Blokland, BI, Theoharis, T. An indexing scheme and descriptor for 3d object retrieval based on local shape querying. Computers & Graphics 2020;92:55–66.
[41] Zhu, A, Yang, J, Zhao, W, Cao, Z. Lrf-net: Learning local reference frames for 3D local shape description and matching. Sensors 2020;20:1–18.
[42] Deng, H, Birdal, T, Ilic, S. Ppfnet: Global context aware local features for robust 3D point matching. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. 2018, p. 195–205.
[43] Ao, S, Hu, Q, Yang, B, Markham, A, Guo, Y. Spinnet: Learning a general surface descriptor for 3D point cloud registration. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021,.
[44] Rusu, RB, Cousins, S. 3D is here: Point cloud library (pcl). In: Proc. IEEE International Conference on Robotics and Automation. 2011, p. 1–4.
[45] Curless, B, Levoy, M. A volumetric method for building complex models from range images. In: Proc. Annual Conference on Computer Graphics and Interactive Techniques. 1996, p. 303–312.
[46] Besl, PJ, McKay, ND. A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 1992;14:239–256.
[47] Guo, Y, Bennamoun, M, Sohel, F, Lu, M, Wan, J, Kwok, NM. A comprehensive performance evaluation of 3D local feature descriptors. International Journal of Computer Vision 2016;116:66–89.
[48] Lu, X, Jain, AK, Colbry, D. Matching 2.5d face scans to 3D models. 2006.
[49] Kordelas, G, Mademlis, A, Daras, P, Strintzis, MG. Object recognition and pose estimation from 2.5d scenes. In: Encyclopedia of Multimedia. 2008, p. 674–676.
[50] Fischler, MA, Bolles, RC. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. ACM 1981;24:381–395.