A Novel SDASS Descriptor for Fully Encoding the Information of 3D Local Surface

Bao Zhao, Xinyi Le, Juntong Xi

Abstract—Local feature description is a fundamental yet challenging task in 3D computer vision. This paper proposes a novel descriptor, named Statistic of Deviation Angles on Subdivided Space (SDASS), for comprehensive encoding geometrical and spatial information of local surface on Local Reference Axis (LRA). The SDASS descriptor is generated by one geometrical feature and two spatial features. Considering that surface normals, which are usually used for encoding geometrical information of local surface, are vulnerable to various nuisances, we propose a robust geometrical attribute, called Local Principal Axis (LPA), to replace the normals for generating the geometrical feature of our SDASS descriptor. For accurately encoding spatial information, we use two spatial features for fully encoding the spatial information of a local surface based on LRA. Besides, an improved LRA is proposed for increasing the robustness of our SDASS to noise and varying mesh resolutions. The performance of the SDASS descriptor is rigorously tested on several popular datasets. Results show that our descriptor has a high descriptiveness and strong robustness, and its performance outperform existing algorithms by a large margin. Finally, the proposed descriptor is applied to 3D registration. The accurate result further confirms the effectiveness of the SDASS method.

Index Terms—Local feature descriptor, local reference axis, object recognition, 3D registration.

1 INTRODUCTION

Local feature descriptor being used for encoding the information of local surface has many applications in 3D computer vision areas such as 3D registration [1-3], 3D object categorization and recognition [4-6], 3D model retrieval and shape analysis [7, 8], and 3D biometric [9], to name a few. In the last few years, with the development of numerous low-cost scanners and high-speed computing systems (e.g., Microsoft Kinect and Intel RealSense), 3D data (e.g., clouds, meshes and depth images) becomes easily available, which further improves the significance of investigating local shape descriptors in 3D computer vision area.

A local shape descriptor is usually constructed by transforming the geometrical and spatial information of a local surface into a feature vector representation [10]. It is worth noting that the descriptor presented in this paper is applied to rigid objects. Therefore, the fundamental attribute of local shape descriptor should be invariant to rigid transformation. Furthermore, a shape descriptor should have a high descriptiveness and robustness [11]. The descriptiveness of a local shape descriptor is an ability of encoding the predominant information on the underlying local surface. In other words, the descriptiveness denotes the ability of distinguishing one local surface from another. The robustness of a local feature descriptor indicates an ability of resisting the impact of various nuisances including noise, varying mesh resolutions, etc. Besides, the compactness and efficiency are also important to a feature descriptor for some applications including robots and mobile phones [12]. So, designing a local feature descriptor with overall good performance for dealing with above mentioned nuisances is a tremendous challenge. Over the last two decades, a number of local feature descriptors are designed for improving the ability of coping with these nuisances. Examples include Spin Images [13, 14], signature of histograms of orientations (SHOT) [10, 15], rotational projection statistics (RoPS) [16]. For more details, readers can refer to a recent survey [11]. For a local feature descriptor, local frame and feature representation are two major elements for determining its performance. The local frame can be divided into two categories. One is defined as local reference frame (LRF), and another is defined as local reference axis (LRA). The LRF is composed of three orthogonal axes, and the LRA only comprises a single orientated axis. Therefore, LRF can provide entire local 3D spatial information including radial, azimuth and elevation directions, while LRA only provide the spatial information in radial and elevation directions. It is worth noting that some local descriptors (e.g. THRIFT [17], PFH [18]) are not using local frame for constructing their features. These descriptors can encode the spatial information only in radial direction, which usually present a limiting descriptiveness owing to the lack of adequate spatial information [11]. For the LRF/A-based descriptors, although LRF can provide entire spatial information, the repeatability of x/y axis are more vulnerable to nuisances (e.g. noise, varying mesh resolution and symmetrical surface) than z axis, and more time is needed for constructing it than LRA [12, 19]. Therefore, a local descriptor constructed on LRA has a high potential with robust to various nuisances [12]. For LRA-based descriptors, the repeatability of LRA directly influence their performance. To achieve a high repeatability of LRA, many methods have proposed for constructing LRA or LRF (the z-axis of a LRF can be used as a LRA). These methods include the techniques proposed by Mian et al. [20], Tombari et al. [15], etc. In these methods, LRA or the z-axis of LRF is usually constructed by covariane analysis. Unfortunately, none of these methods have an overall good performance for coping with various nuisances including noise, mesh boundary, sign ambiguity, etc.

In the view of the feature representation of a descriptor, the geometrical and spatial information of a local surface are usually...
encoded for representing a local shape. Some existing descriptors only encode the geometrical information of a local surface [17, 21]. In these methods, the deviation angle between normals or between normal and LRA is a popular way of encoding local geometrical information. However, owing to low repeatability of normal, geometrical information of local surface encoded by the deviation angle presents a lower robustness [11]. Therefore, these descriptors of encoding only geometrical information often present a poor performance for resisting noise, varying mesh resolution, etc. In contrast, some descriptors only encode the spatial information of a local surface [13, 16, 22]. However, some of these descriptors are incomplete to encode the spatial information by transforming 3D to 2D/1D (e.g. RoPS [16], TriSI [22]), and some of these descriptors are redundant to encode the spatial information by repetitively using the information of x, y and z coordinates of local points (e.g. RoPS [22], TOLDI [23]). In addition, some descriptors encode both geometrical and spatial information of a local surface [3, 15]. Although, encoding geometrical information together with spatial information will significantly improve the descriptiveness of a feature descriptor [15], as mentioned above, lower robustness of geometrical feature and the imperfect encoding spatial information also limit the descriptiveness and robustness of these descriptors.

In these regards, we propose a novel local feature descriptor named Statistic of Deviation Angles on Subdivided Space (SDASS). Our SDASS is generated on LRA and encode both geometrical and spatial information of a local surface. Specifically, we first propose an improved LRA which is developed from the LRF proposed by Yang et al. [23]. Considering normal being vulnerable to various nuisances, to improve the robustness of encoding geometrical information, we propose a new geometrical attribute named local principal axis (LPA), which is verified of having a strong robustness to resist various nuisances, to replace normals for constructing the deviation angle between LPA and LRA. For encoding spatial information, our SDASS use two spatial features for fully encoding spatial information of a local surface on LRA. Different from some previous descriptors which need to process initial point cloud such as triangulation [16, 22], our SDASS is directly performed on the initial point clouds. For verifying the performance of the SDASS, it is applied to three popular datasets including shape retrieval, shape registration and object recognition scenarios, and compared to several state-of-the-art methods. The experimental results show that the performance of our SDASS exceeds the existing methods by a large margin. In addition, the SDASS is applied to 3D registration in the last of this paper. The accurate outcomes further confirm the effectiveness of our method.

The major contributions of this paper are summarized as follows: First, we propose a geometrical attribute LPA which has a significantly high repeatability compared to normals. We use the proposed LPA to replace normals for generating deviation angles, which presents a high descriptiveness and strong robustness for encoding the geometrical information of local surface. Second, a novel local shape descriptor named SDASS is proposed. The SDASS is generated on LRA, and describe a local surface by the combination of encoding geometrical and spatial information. The experimental results show that the performance of SDASS significantly surpasses the existing local shape descriptors. Third, an improved LRA is proposed, which has a superior performance for resisting noise and varying mesh resolution compared to the existing LRF/A.

The rest of this paper is organized as follows. Section 2 provides a brief review of related work of generating LRF/A construction and local shape descriptors. Section 3 presents a detailed description of the proposed SDASS method. Section 4 introduces the experimental evaluation of our method and several state-of-the-art methods on three standard datasets. Section 5 describes the simple application of the SDASS descriptor on 3D registration. The conclusions and future works are drawn in Section 6.

2 RELATED WORK

This section presents a brief overview of the existing methods for local surface feature description. Considering that the SDASS method belong to the family of LRA-based methods. The existing methods for constructing LRA or the z-axis of LRF are first reviewed. Then, the existing local shape descriptors are divided into three categories to be described respectively.

2.1 The Methods of Constructing LRA or the Z-Axis of LRF

Most existing methods of constructing LRA or the z-axis of LRF are based on covariance [19]. In these methods, the z-axis or LRA is the normalized eigenvector corresponding to the minimal eigenvalue of a covariance matrix. Mian et al. [20] use radius neighbors instead of k nearest neighbors of a key point to construct the covariance matrix for improving the robustness to varying mesh resolution. However, the sign of the LRF is not defined unambiguously. Tombari et al. [10] construct a weighted covariance matrix by first using a key point to replace the barycenter of radius neighbors, and then assigning smaller weights to more distant points. The sign of this LRF is unambiguous by inclining the barycenter of local surface. This method has been proven to be quite robust to noise, while is vulnerable to varying mesh resolution [23]. Later, Guo et al. [16] propose a novel technique for constructing LRF by first applying two weighting strategies to each triangle on a local surface, and then using all weighted triangles to calculate the covariance matrix. The sign of this LRF is disambiguated by aligning the direction to the majority of the point scatter. The main advantage of this method is to have a high robustness to varying mesh resolution [16], while the efficiency is very low [23]. Recently, Petrelli et al. [1] and Yang et al. [23] pick a small subset of the radius neighbors for constructing covariance matrix. The main purpose of using less points for calculating the z-axis is to improve its robustness to occlusion, clutter and mesh boundaries. In addition, Andrew E. Johnson et al. [13] and Yang et al. [3] directly use normal as LRA. Since normal is vulnerable to various nuisances (e.g. noise, varying mesh resolution), the LRA present a low robustness.

Some methods mentioned above have nicely solved the problem of sign ambiguous, and are robust to some nuisances. However, they do not have an overall good performance for coping with all nuisances (e.g. noise, varying mesh resolution, etc.), such as the LRF proposed by Guo et al. [16] has a strong robustness to varying mesh resolution and noise while it is vulnerable to mesh boundary, and the LRF proposed by Yang et al. [23] is robust to mesh boundary, while it is vulnerable to noise and varying mesh resolutions.
2.2 Local Shape Descriptors

Local shape descriptors have been widely proposed in literatures over the last two decades [11, 24]. Among these descriptors, some only encode geometrical information, and some only encode spatial information, and others encode the combination of geometrical and spatial information of a local surface.

There are some descriptors for only encoding geometrical information. Flint et al. [17, 25] proposed a local feature descriptor called THRIFT which uses the deviation angles between the normal at a key point and the normals at its neighbors to construct a 1D histogram. The THRIFT is a very efficient descriptor, while it is very sensitive to noise and varying mesh resolutions [11]. Rusu et al. [18] proposed a point feature histogram (PFH) by using several features of point pairs in the support region. The several features are obtained on a Darboux frame constructed by the normals and point positions. The PFH is robust to varying mesh resolutions, while it is vulnerable to noise [11]. Later on, in order to improve the efficiency, they used the simplified point feature histogram (SPFH) of neighbors to construct the fast point feature histograms (FPFH) descriptor [21] which has a similar performance with PFH in descriptiveness and robustness.

Some descriptors only encode spatial information. Johnson and Hebert [13] proposed a local shape descriptor named Spin Image (SI). The Spin Image use the normal at a key point as the LRA, and then each point in the support region is represented by two spatial distances. Finally, the Spin Image is generated by accumulating the number of local points into each bin of the 2D array. The Spin Image completely encodes the spatial information of a local surface on LRA while it is sensitive to noise [11] owing to the low repeatability of its LRA. Tombari et al. [26] proposed a unique shape context (USC) which is developed from 3DSC descriptor [27]. The USC is generated on a LRF, proposed by Tombari et al. [10], by dividing the local 3D space into bins along azimuth, elevation and radial directions. The USC completely encode the spatial information on LRF, and present a high robustness to noise, clutter and occlusion, while it is sensitive to varying mesh resolutions [11]. Guo et al. [16, 28] proposed the rotational projection statistics (RoPS) descriptor. In RoPS descriptor, a novel LRF is first constructed, and then the feature representation is generated by calculating a set of statistics of point density with respect to numerous rotations of local surface around each axis. The RoPS descriptor is proved to have a high descriptiveness [11], while it is very time-consuming. Similar to the view-based mechanism in RoPS, Guo et al. [22, 29] proposed a Tri-Spin-Image (TriSI). TriSI is also generated on a LRF which constructed by a similar technique as in the RoPS. Then, TriSI is generated by integrating and compressing three spin image signatures created on each axis of the LRF. The performance of TriSI to resist various nuisances (e.g. noise, varying mesh resolutions) is slightly better than that of the RoPS [11]. Like RoPS, TriSI is also time-consuming. Recently, Yang et al. [23] proposed a novel triple orthogonal local depth images (TOLDI) by first constructing a LRF, and then concatenating three local depth images captured from three orthogonal view planes in the LRF to generate a feature vector. Although TOLDI achieve a good performance for resisting various nuisances, the feature vector of TOLDI present a low compactness [22].

There are also some descriptors for encoding the combination of geometrical and spatial information. Tombari et al. [10, 15] proposed the signature of histograms of orientations (SHOT) descriptor based on a LRF. The SHOT descriptor first encodes spatial information on the LRF by dividing the spherical neighborhood space into several bins along the radial, azimuth and elevation directions. Then, for each bin, the geometrical information is encoded by generating the deviation angles between the normal at the key point and the normals at the radius neighbors. Despite SHOT having a high descriptiveness, it is sensitive to varying mesh resolutions [11, 23]. Recently, Yang et al. [3] proposed a local feature statistics histograms (LFSH) by concatenating three features, including two spatial features (local depth and horizontal projection distance) and one geometrical feature (deviation angle), into a feature vector. LFSH is very compact and efficient, while it suffers from limited descriptiveness [3].

In conclusion, encoding geometrical information combined with spatial information can improve the descriptiveness of a descriptor, while some existing descriptors only encoding geometrical or spatial information, which have limited descriptiveness. In addition, for encoding geometrical information, the deviation angles between normals or between normal and LRA is commonly used. However, the normal generated on a very small local region [30] has a weak robustness to resist various nuisances. For encoding spatial information, some descriptors (e.g. RoPS, TriSI) don’t fully encode the spatial information on LRF by transforming 3D to 2D/1D, and some others aren’t compact to encode spatial information by repetitive using the information of x, y and z coordinates of local points (e.g. TriSI, TOLDI). Therefore, the descriptiveness and robustness of existing descriptors have the potential for further being improved.

3 SDASS-BASED LOCAL SHAPE DESCRIPTION

This section details our SDASS technique for local shape description. Specifically, we first introduce an improved LRA and robust LPA. Then, we present the SDASS feature representation by generating three features (projected radial distance, height distance and deviation angle between LPA and LRA) based on the proposed LRA and LPA. Finally, the key parameters of SDASS are quantitatively selected.

3.1 Constructing Local Reference Axis (LRA) and Local Principal Axis (LPA)

In this section, we construct a LRA developed from the z-axis in the LRF proposed by Yang et al. [23]. To improve the robustness and descriptiveness of encoding geometrical information, we define a new geometrical attribute named local principal axis (LPA) to replace normal for calculating the deviation angles.

3.1.1 Local Reference Axis

In this paper, we select LRA rather than LRF as a basis for spatial division. Although, LRF provides the entire local 3D spatial information including radial, azimuth and elevation while LRA loses one-dimension spatial information (i.e., the azimuth information), as shown in Fig.1. However, the repeatability and robustness of the x, y axis are significantly lower than the z axis in a LRF [12, 19], especially in the presence of noise, varying mesh resolution, etc. To provide a more accurate spatial division in our feature descriptor, we select LRA as a frame to construct our descriptor.

Our LRA is developed from the z-axis in the LRF proposed by Yang et.al. [23]. In below, we first simply repeat the z-axis proposed by Yang et.al. [23], and then propose our improved
Given a key point $p$ and a support radius $R$, local points $Q = \{q_1, q_2, ..., q_n\}$ can be obtained within the distance of $R$ from the key point $p$. The subset of $Q$, which is denoted as $Q_z = \{q_1^z, q_2^z, ..., q_n^z\}$, determined by $1/3$ support radius neighbors around the key point is used for calculating the direction of the $z$-axis. Specifically, based on $Q_z$, a covariance matrix $\text{Cov}(Q_z)$ is constructed as:

$$\text{Cov}(Q_z) = \begin{bmatrix} q_1^z - \bar{q}^z & \cdot & \cdot & \cdot \\ \cdot & q_2^z - \bar{q}^z & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & q_n^z - \bar{q}^z \end{bmatrix}$$ (1)

where $n$ is the size of $Q_z$, and $\bar{q}^z$ is the centroid of $Q_z$. The eigenvector $\mathbf{v}(p)$ corresponding to the minimum eigenvalue of $\text{Cov}(Q_z)$ is computed. Next, all support radius neighbors are used for disambiguating the sign of $\mathbf{v}(p)$ as follow:

$$z(p) = \begin{cases} \mathbf{v}(p), & \text{if } \mathbf{v}(p) \cdot \sum_{i=1}^{n} q_i \mathbf{p} \geq 0 \\ -\mathbf{v}(p), & \text{otherwise} \end{cases}$$ (2)

First, in the presence of noise and varying mesh resolutions (as shown in Fig.2 (a)), the repeatability of the LRA generated by $S1$ and $S2$ is gradually improved along with the increase of support radius. The highest repeatability values of them are the same appeared with the support radius at 20mr, although the repeatability of the LRA generated by $S1$ is higher than $S2$ in the process of the support radius increased from 3mr to 20mr. So, we can conclude that, for the impact of noise and varying mesh resolution, the support radii for calculating the direction and eliminating the sign ambiguity by simultaneously taking the maximum value (i.e. 20mr) can get a higher repeatability.

Second, for the impact of mesh boundary as tested on UWAOIR dataset in Fig.2. (b), the repeatability of the LRA generated by $S1$ is higher than $S2$, and this difference gradually disappear along with the increase of support radius. In addition, the varying tendency of them is similar. Their repeatability is gradually improved along with the increase of the support radius before about 6mr, and then drop after 6mr. When eliminate the influence of mesh boundary, as tested on UWAOIR-IR dataset in Fig.2 (b), the repeatability of the LRAs constructed by $S1$ and $S2$ isn’t fall along with the increase of support radius. From this observation, we can conclude that the strategy $S1$ has a superior performance for resisting the impact of mesh boundary compared to $S2$, while, when the mesh boundary is eliminated (as tested on UWAOIR-IR), the best performances of $S1$ and $S2$ are similar in the process of support radius from 3mr to 20mr.
From the above two observations and corresponding conclusions, we can find that our LRA has a strong robustness to noise and varying mesh resolution, while has a weak robustness to mesh boundary compared to the z-axis proposed by Yang et al. [23]. The reason of improving the robustness to noise and varying mesh resolution and sacrificing the robustness to mesh boundary include two aspects. The first is that, for a LRA, it is difficult to have a comprehensively good performance for resisting various nuisances (e.g. noise, varying mesh resolution and mesh boundary, et al.). The second is that mesh boundary is easier eliminated than noise and varying mesh resolutions from a point cloud [11, 20]. So, the improved LRA is appropriate applied on a point cloud with no boundary existent (e.g. the BR dataset) or boundary identified (e.g. the UWAOR-IR dataset). In the case of mesh boundary being existent and not identified, the z-axis of the LRFs (e.g. the LRF proposed by Yang et. al. [23]) with robust to mesh boundary can be used as the LRA to generate our descriptor.

3.1.2 Local Principal Axis

The deviation angle between normals or between normal and LRA is commonly used for encoding geometrical information in previous descriptors (e.g. FPFH [21], SHOT [15]). In this process, the repeatability of normals determines the performance of encoding geometrical information on a local surface. At present, there are two popular techniques of generating normal. The first technique uses all triangular patches attached with a point to calculate the normal of that point [11, 20, 32]. The second technique first generating a covariance matrix by using the radius neighbors of a point, and then calculating the eigenvector corresponding to the minimum eigenvalue as for the normal of that point [30, 33, 34]. In the following, for increasing readability, the normal generated by the first method is called triangular patches based normal (TN) and by the second method is called radius neighbors based normal (RN). However, the disambiguation of the normals constructed by these two techniques rely on a viewpoint [13, 34]. If the viewpoint is unknown, manual to disambiguate of normals is needed [34], which is inefficiency.

In addition, normal is a geometrical attribute to represent a very small local region (e.g. a local surface for generating TN has a support radius of 1mr averagely, and a local surface for constructing RN usually has a support radius less than 3mr [30, 35]). The repeatability of the normal generated on a small local region is sensitive to noise, varying mesh resolution, etc. (as presented in Sect.3.1.1).

For overcoming the above two weaknesses existing in normal, we here propose a local geometrical attribute called Local Principal Axis (LPA) to replace normal for constructing deviation angle between LRA and LPA. The method of generating our LPA is similar to the technique of generating the improved LRA introduced in Sect.3.1.1. In opposite to use a small region for calculating the direction of normal, the proposed LPA use a larger local region for determining its direction, which significantly increase the robustness to various nuisances. According to the tested results presented in Fig.2, considering the tradeoff among efficiency, the robustness to noise and varying mesh resolution and the robustness to mesh boundary, the support radius for generating LPA is selected as 7mr in this paper. Based on this support radius, the direction and sign of LPA can be determined by using Eqs. (1) and (2), respectively, which are presented in Sect.3.1.1. In this way, disambiguating the sign no longer rely on the view point for improving the efficiency and accuracy.

3.2 SDASS Descriptor

Once the LRA and LPA are constructed, we are left with the task of encoding spatial and geometrical information contained on a local surface. Given a point cloud or surface, the local points are determined by a key point \( p \) and a support radius \( R \). We define the local points as \( \mathbf{Q} = \{ q_1, q_2, \ldots, q_m \} \). As shown in Fig. 4 (b), (c), \( \mathbf{Q} \) is first transformed to make \( p \) coincided with the coordinate origin, and the LRA of the key point aligned with \( z \) axis to achieve the rotation invariance of the local surface. The transformed local points are denoted as \( \mathbf{Q}' = \{ q_{1}', q_{2}', \ldots, q_{m}' \} \). Then, we seek for an appropriate manner of encoding spatial and geometrical information on the local surface.

![Fig. 3. Two ways of encoding spatial information on LRA](image)

In general, the spatial information of local surface on LRA can be fully encoded by two ways, as shown in Fig. 3. One way uses the radial distance (\( \rho \)) and azimuth angle (\( \theta \)) to encode spatial information, and another way use the height distance (\( h \)) and projected radial distance (\( d \)) to encode spatial information. Some descriptors use the radial distance and azimuth angle to encode spatial information (e.g. SHOT [15], USC [26]), while the majority of descriptors select to use the height and projected radial (e.g. Spin image [13], TrSi [29], LFSH [3]). In this paper, we also use the projected radial and height distance for comprehensive encoding the spatial information on LRA. Besides, our SDASS descriptor uses a geometrical feature (the deviation angle between LRA and LPA) to encode the geometrical information of local surface. The deviation angles between LRA and the normals of neighbors are usually used for encoding geometrical information on local surface, such as SHOT [15], LFSH [3]. Considering that normal represents the attribute of small local region, which is sensitive to various nuisances, we propose a geometrical attribute LPA (see Sect. 3.1.2), which has very high robustness compared to normals as verified in Sect. 4.2.1, to replace the normal for constructing the deviation angles. Next, we present the process of generating the three local features of the SDASS descriptor.

1. Projected radial distance information, as shown in Fig.4 (d):

   For the transformed local points \( \mathbf{Q}' \), the tangent plane at the key point \( p \) is coincided with XY plane. The \( \mathbf{Q}' \) are projected onto the tangent plane for obtaining 2D projected points. The distances between these 2D points to the \( p \) are computed. Since the LRA has aligned with z axis, the distances are simply computed as:

   \[ l_i^p = \sqrt{q_i'^{(1)}^2+q_i'^{(2)}^2} \quad i \in 1, 2, \ldots, m \]

   where \( q_i'^{(1)} \) and \( q_i'^{(2)} \) denote the x, y coordinate of the point \( q_i' \), respectively. The range of \( l_i^p \in [0, R] \).

2. Height distance information, as shown in Fig.4 (e):

   We translate the tangent plane with a distance of \( R \) along the negative direction of LRA. The height distances are defined as the distances between the local points \( \mathbf{Q}' \) and the translated tangent plane, which are calculated as:
The deviation angles between LRA and LPA, as shown in Fig. 4 (f):

The deviation angles between the LRA of the key point \( p \) and the LPA of the local points \( \mathbf{Q}' \) are used for encoding geometrical information. Similar to normals, the LPA of all points on the point cloud is needed to be precomputed. Based on the precomputed LPA of the local points, the deviation angles corresponding to this key point are calculated as:

\[
\theta_i = \arccos(LRA, LPA) \quad i \in 1, 2, ..., m
\]

where \( \theta_i \in [0, \pi] \).

First, in opposite to use the deviation angle of normals in previous descriptors (i.e. LFSH [3], FPFH [21]), the SDASS use the deviation angle between LPA and LRA for encoding geometrical information of local surface. The LPA has a high repeatability in the presence of various nuisances (e.g. noise, varying mesh resolutions, etc.) compared to normal (as verified in Sect. 4.2.1). The repeatability of LPA/normal determines the accuracy of calculating the deviation angle. So, with the high repeatability of our LPA, the geometrical information encoded in our SDASS descriptor has a strong robustness and a high descriptiveness.

Second, the SDASS descriptor encode the combination of spatial and geometrical information of a local surface. The spatial information on LRA is completely encoded by two features (projected radial distance and height distance). The geometrical information is encoded by the deviation angles between LRA and LPA. In the view of the components in SDASS descriptor, LFSH is the most similar descriptor to ours. In contrast, the three features used in LFSH are concatenated to constitute the final 1D histogram. Although, it can reduce the dimensions of the descriptor, it is unable to fully encode the spatial information on LRA, and partitioned statistics of the geometrical information. In addition, the LFSH directly use the normal at a key point as LRA and encodes local geometrical information by using the deviation angles between the LRA and the normals of neighbors, which greatly reduces the robustness of descriptor to various nuisances (e.g. noise, varying mesh resolution, etc.) owing to the low repeatability of normal (as verified in Sect. 4.2.1).

Third, the SDASS descriptor is generated on LRA which provide more accurate spatial information than LRF [12, 19]. We
propose an improved LRA for generating SDASS descriptor. The LRA has a strong robustness to noise and varying mesh resolution while it is sensitive to the boundary of points cloud (as verified in Sect. 4.2.1). The reason of generating this LRA is of considering that mesh boundary is easier eliminated than noise and varying mesh resolutions from a point cloud [11]. So, we can select to use the improved LRA in the case of a point cloud with no boundary existent or boundary identified. In the case of a point cloud having boundary and not identified, z-axis of the LRFs (e.g. the LRF proposed by Yang et. al. [22]) with robust to mesh boundary can be used as the LRA in our descriptor. So, a strong robustness of LRA can guarantee a good performance of the SDASS descriptor.

Fourth, the computational efficiency of our SDASS descriptor is high. The computational process of the SDASS descriptor mainly include three steps: transforming local points; computing the three features; mapping local points from 3D geometrical space to 3D feature space. Obviously, these steps are all simple arithmetic. The computational complexity of each step is O(k), where k denotes the number of the local points. So, the computational complexity of SDASS descriptor is also O(k).

In addition, unlike to some descriptors (e.g. RoPS [16], TriSI [22]) need the mesh information of points cloud, the SDASS descriptor is directly applied on original points cloud.

3.3 Selecting SDASS Parameters

There are four parameters to generate our SDASS descriptor. They are respectively the support radius R and the number of subdivisions of the three features \( N_r, N_i, \) and \( N_o \). The support radius R is an important parameter because large values of R would affect the computational efficiency and increase the descriptor’s sensitivity to mesh boundary, clutter and occlusion, whereas small values of R would make a descriptor less distinctive [16]. According to the suggestion in [12, 31], we select 20 mesh resolution (hereinafter \( mR \), computed as either the average length of the edges of the meshes or, should the dataset consist of point clouds, the average distance between neighbor points [1]. In this paper, we consistently use the second method to calculate \( mR \) as the value of support radius R. In this paper, the dimension of SDASS descriptor is \( N_r \times N_i \times N_o \). Obviously, the big values of \( N_r, N_i, N_o \) will affect the computational efficiency and consume more memory. If the values are too small, the descriptor would lose many details about local shape. According to our experimentation and refer to the setting in [3, 15], the three parameters \( N_r, N_i, N_o \) are set to 5, 5, 15 in this paper, respectively.

4 Experiments

In this section, our LRA, LPA and SDASS descriptor are tested on three standard datasets for verifying the robustness to noise, varying mesh resolution, mesh boundary, and etc. The three datasets include the Bologna retrieval (BR) dataset [36, 37], the UWA object recognition (UWAOR) dataset [6], and the UWA 3D modeling (UWA3M) dataset [6, 38]. Our method is also compared with several state-of-the-art methods (including Spin Image [13], SHOT [15], RoPS [16], TriSI [22], LFSH [3], TOLDI [23]) for contrastive evaluation. All tested descriptors are implemented in MATLAB. The MATLAB code of SHOT is provided by author and the MATLAB code of Spin Image, RoPS and TriSI are available in website\(^1\). The MATLAB codes of LFSH is written by us from their corresponding C++ versions which is provided by author. The MATLAB code of TOLDI is written by us of referring to the author’s paper [23]. All the experiments in this paper are implemented on a PC with an Intel Core i7-4790 CPU and 12GB of RAM.

4.1 Experimental Setup

In the following, the implementation details, including the description of datasets and the adopted evaluation criteria, of the experiments are introduced.

4.1.1 Datasets

Like [23], considering the BR, UWAOR and UWA3M datasets can cover various nuisances (e.g. noise, varying mesh resolution, clutter and occlusion, etc.), we also use these three datasets in this paper for comprehensive evaluating the performance of our descriptor. Some exemplars of the three datasets are shown in Fig.5. Specifically, the BR datasets contains noise, and the UWAOR dataset contains occlusions, clutter and mesh boundary, and the UWA3M dataset contains occlusions and mesh boundary. In addition, for comprehensive verifying the robustness to noise and varying mesh resolutions, we generate some new scenes in the BR dataset, and for presenting the influence of mesh boundary, we generate two new datasets developed from UWAOR and UWA3M datasets, respectively. The details are presented as follows.

The BR dataset has 6 models and 18 synthetic scenes. The six models are noise freely copied from the Stanford 3D Scanning Repository [36], which are scanned by a Cyberware 3030 MS scanner. The 18 scenes in the BR datasets are created by adding Gaussian noises with the standard deviation of 0.1, 0.3, and 0.5 \( mR \), respectively, in each randomly transformed model. To give a comprehensive comparison for the robustness of descriptors to noise and varying mesh resolution, three group scenes are generated from the BR dataset. Each group includes 60 scenes. The first group is generated by adding Gaussian noises with increasing standard deviations from 0.1 to 1.0 \( mR \) with an interval of 0.1 \( mR \) in each randomly transformed BR model, which include the 18 scenes copied from original BR dataset. The second group is created by resampling each randomly transformed model from 10/10 (10/10 is the original resolution) to 1/10 of their original mesh resolution with an interval of 1/10. The third group is generated by correspondingly combining the all levels of Gaussian noise in the first group with the all levels varying mesh resolutions in the second group (e.g. the standard deviation of 0.1 \( mR \) combined with the decimation rate of 10/10). The scenes in this dataset without any occlusion or clutter. The purpose of employing this dataset is to verify the robustness to noise and varying mesh resolutions.

The UWAOR dataset contains 5 models and 50 real scenes. The scenes are generated by randomly placing four or five models together in a scene and scanned from a specific viewpoint using a Minolta Vivid 910 scanner. Mesh boundary, clutter and occlusion are the major challenges in this dataset. The UWA3M dataset contains 22, 16, 16, and 21 2.5D views scanned respectively from four objects by using a Minolta Vivid 910 scanner.

---

\(^1\) The website of Spin Image is: http://staff-home.ecm.uwa.edu.au/~00053650/code.html, and the website of RoPS and TriSI is: http://yulanguo.me/.
Due to that a single viewpoint cannot capture the structure of an entire 3D model, feature description in this dataset are confronted with the nuisances including mesh boundary and self-occlusion. Since every two views of an object in the UWA3M dataset do not always share an overlap. To guarantee model-scene views of an object having an overlap, five pairs of views from each object with bigger overlap area are selected to test our descriptor (20 view pairs in total).

![Fig. 5. One exemplar model and One corresponding exemplar scene (shown from left to right) respectively taken from the BR, UWAOR and UWA3M datasets.](image)

In addition, to present the influence of mesh boundary, we generate two new datasets derived from UWAOR and UWA3M, respectively. For increasing readability, the two new datasets are named UWAO-IR and UWAO-IR, respectively. The new dataset has the same models and scenes with their original dataset. The only difference between them is that the inner and boundary region of the scenes in UWAO-IR and the scenes and models in UWA3M-IR need to be distinguished. The inner region in a point cloud is defined as the region in which points have a distance larger than the support radius R from the mesh boundary. One exemplar of scene in UWAO-IR and scene or model UWA3M-IR are shown in Fig. 6. The purpose of distinguishing inner and boundary region is to avoid the influence of mesh boundary by sampling key points only on the inner region, which will be detailed in Sect 4.1.2.

4.1.2 Evaluation Criteria

We use the precision-recall curve (PRC) to evaluate the performance of feature descriptors, and directly use angle error to present the repeatability of the axes in LRF, LRA, LPA and normals (TN, RN).

To verify the repeatability of proposed LRA and LPA, the angle error between two axes is used as an evaluation criterion. The angle error of two arbitrary 3D axes $v_1$ and $v_2$ can be simply computed as follow.

$$e = \arccos\left(\frac{v_1 \cdot v_2}{||v_1|| ||v_2||}\right)$$

The PRC is a popular method for evaluating local feature descriptors. The detailed process of generating PRC can refer to [11, 16, 23]. If the descriptor feature obtains both high recall and precision, the PRC would fall in the top left of the plot. In order to compactly and quantitatively present the performance of feature descriptors, the area under the PRC curve, defined as AUCpr, is also used in this paper. AUCpr is a simple and aggregated metric to measure how an algorithm performs over the whole precision-recall space [11]. In the ideal case, the AUCpr is equal to 1 since the recall is always 1 for any precision.

In our experiments, for the BR and UWAOR datasets, 1000 key points are randomly selected from scene, and then their corresponding key points are extracted from the models/model using the ground truth transformation which are given by the publishers. For the UWA3M dataset, each pair of model-scene has an overlapping region. The overlapping region is determined based upon the ground truth transformation which is obtained by first manually aligning the model-scene pair and then refining using the ICP algorithm [39, 40]. Then, 1000 key points are randomly selected from the overlapping region of the scene, and their corresponding key points are extracted from the model using the ground truth transformation. For UWAOR-IR dataset, the process of selecting key points is similar to that of UWAOR except that the 1000 key points on scenes are randomly selected from their inner region. For UWA3M-IR dataset, the overlapping inner regions on each model-scene pair is constructed. The key points in UWA3M-IR are selected on the overlapping inner regions by the similar method of selecting key points in UWA3M.

After the key points are determined, if we only evaluate the repeatability of LRF/A, LPA and normals, we just need to generate LRF/A, LPA and normals on these key points, and use the angle error to evaluate them. If we evaluate the performance of the local feature descriptors, we need to generate the descriptor features on these key points. The PRC curve or its area AUCpr is used for comparing the performance of the descriptors.

4.2 Performance Evaluation of the Proposed LRA and LPA

In this section, the repeatability and efficiency of our proposed LRA and LPA are tested. The proposed LRA and LPA are compared with five recent techniques of constructing LRFs and two popular methods for generating normal. The five techniques of LRFs include Mian et al. [20], Tombari et al. [10], Petrelli et al. [1], Guo et al. [16] and Yang et al. [22]. The two methods for constructing normals are respectively triangular patches based method (TN) [11, 32] and radius neighbors based method (RN) [30, 33], which have been introduced in Sect. 3.1.2. For convenient expression, LRF, LRA, LPA and normals are collectively called local axes in the following.

4.2.1 Repeatability Performance

Our proposed LRA and LPA are tested on the three datasets, which are detailed in Sect. 4.1.1, for evaluating their repeatability. The support radius of constructing these local axes is kept the same as 20 \text{mr} for a fair comparison. According to the implementation details introduced in Sect. 4.1.2, we can obtain the angle errors of each method tested on the five experimental datasets (i.e. BR, UWAOR, UWA3M, UWAO-IR and UWA3M-IR). The percentage of the angle errors below 5° corresponding to each method is counted, as shown in Fig. 7. In order to compare the repeatability of the x and z axis in a LRF, the angle errors of the x and z axis in a LRF are computed respectively.

In the BR dataset, the robustness of our proposed LRA and LPA to noise and varying mesh resolution are verified. For conciseness, we select a few of scenes from the scenes of BR datasets for this experiment. The results of the local axes tested on
the selected scenes of BR dataset are showed in Fig. 7 (a)-(c). Several observations can be drawn from these figures. First, the z-axis of Guo et al. and our LRA achieve significantly strong robustness to all levels noises and varying mesh resolutions. In particular, the z-axis of Guo et al. achieves the best performance for resisting varying mesh resolutions, as shown in Fig. 7 (b). It may owe to the z-axis proposed by Guo et al. use all the information of local surface rather than only the mesh vertices. Our LRA achieves the best performance in the presence of Gaussian noise and Gaussian noise combined with varying mesh resolution, as shown in Fig. 7 (a), (c). Second, for the five tested LRFs, the performance of the z axes surpasses the corresponding x axes by a large margin. It implies that the repeatability of a LRA can significantly outperforms that of a LRF since the z-axis in a LRF can be as a LRA, which is the major concern of us to use a LRA rather than a LRF for constructing our descriptor. Third, for the results of our LPA and the two normals (i.e. TN and RN), the performance of our LPA significantly outperforms TN and RN in all the cases, and the gap appears to be more obvious along with the increase of the levels of noise and varying mesh solutions. Therefore, our proposed LPA can be integrated in the descriptors (e.g. LFSH, SHOT, etc.), which include the component generated by the normals, for improving their robustness.

In conclusion, our LRA is appropriate to be applied on the dataset with the boundary being identified or no boundary existent. In the presence of the boundary existent in a point cloud, the z axis of some state-of-the-art LRFs (such as Yang et al. [23]) with robust to mesh boundary can be used as the LRA in our descriptor. Our proposed LPA has a strong robustness to various nuisances (e.g. noise, varying mesh resolution, clutter, occlusion, et al.), and significantly outperform the two normals (i.e. TN, RN).

In the UWAOR and UWA3M datasets, some different challenges (e.g. mesh boundary, clutter, occlusion) are existent. Besides, in order to verify the influence of mesh boundary, the datasets UWAOR-IR and UWA3M-IR are used for testing these local axes. The UWAOR-IR and UWA3M-IR are developed from UWAOR and UWA3M, respectively, by identifying their mesh boundary region. The results of testing on the above four datasets are shown in Fig. 7 (d). Several observations can be drawn from this figure. First, our LRA and the z-axes proposed by Yang et al. [23] and Petrelli et al. [1] have a better performance tested on UWAOR and UWA3M datasets, while they superior performances are not existent when tested on UWAOR-IR and UWA3M-IR datasets. The things in common of the above three axes distinguished from the other axes is that the directions of them are calculated by using a subset of support radius region rather than whole support radius region. So, it is obvious that using a smaller radius to calculate a local axis will has a strong robustness to mesh boundary. Second, the performance of each method tested on UWAOR-IR and UWA3M-IR correspondingly outperform that tested on UWAOR and UWA3M. It implies that mesh boundary existing in a point cloud has a great influence to the performance of these local axes. Third, for the comparison of LPA, TN and RN, our LPA outperform TN and RN by a large margin on all the four datasets. Besides, since TN and RN are generated on a small local region, TN and RN have a stronger robustness to resist occlusion, clutter and mesh boundary than to resist noise and varying mesh resolution. Fourth, Our LRA achieves a good performance on UWAOR-IR and UWA3M-IR datasets while an inferior performance on UWAOR and UWA3M datasets. It shows that our LRA is vulnerable to the mesh boundary.

In conclusion, our proposed LRA has a strong robustness to noise and varying mesh resolutions, while has a weak robustness to mesh boundary. So, our LRA is appropriate to be applied on the dataset with the boundary being identified or no boundary existent. In the presence of the boundary existent in a point cloud, the z axis of some state-of-the-art LRFs (such as Yang et al. [23]) with robust to mesh boundary can be used as the LRA in our descriptor. Our proposed LPA has a strong robustness to various nuisances (e.g. noise, varying mesh resolution, clutter, occlusion, et al.), and significantly outperform the two normals (i.e. TN, RN).
4.2.2 Time Efficiency

In this section, the efficiency of the methods for generating these local axes (including the six LRFs, our LRA, our LPA, TN and RN) is tested. Since the efficiency is mainly correlated to the number of points in local region, we only use the BR dataset to test these local axes. In particular, we first randomly select 1000 points in each model on the BR dataset (6000 points in total). For the test of the six LRFs and our LRA, the total time costs of each LRF or our LRA generated on these points with respect to different support radius R are counted. Similar to [23], R also increases from 5 mr to 40 mr with an interval of 5 mr in this paper. For the test of TN, RN and our LPA, since they denote the local geometrical attribute of a surface, we count the total time costs of RN and our LPA on the selected points with a fixed support radius, and the total time cost of PN on the selected points based on the triangular patches attached with these points. The fixed support radius of RN and our LPA are set to 3mr and 7mr, respectively.

![Time Efficiency of the five LRF and our LRA.](image)

The results of the LRFs and our LRA are shown in Fig.8. It can be observed that the computational time of our LRA and the LRFs proposed by Mian et al. [20] and Tombari et al. [10] are similar, and they achieve the best performance in terms of computational efficiency compared to others. The LRF proposed by Guo et al. [16] is significantly inferior to the others in terms of computational efficiency, which is because the LRF is generated based on a local surface rather than the vertices of the local surface. The LRFs proposed by Yang et al. [23] and Petrelli et al. [1] achieve a medium performance in terms of time efficiency. For the comparison of TN, RN and our LPA, the computational times of TN, RN and our LPA are 0.0924s, 0.2820s and 0.3618s, respectively. We can see that the computational efficiency of TN outperforms RN and LPA, while the computational efficiency of TN and our LPA are comparable. It is mainly because that TN and our LPA need to calculate a covariance matrix while TN do not need to. Although our LAP has the worst performance in terms of computational efficiency, the gaps of the computational time among them are not obvious. Besides, our LAP is directly generated on a point cloud, while constructing the TN need a point cloud with the information of triangular mesh.

4.3 Performance Evaluation of the SDASS

In this section, the proposed SDASS descriptor is tested on the five experimental datasets (i.e. BR, UWAOR, UWA3M, UWAOR-IR and UWA3M-IR) using the PRC curve and AUCpr (introduced in Sect. 4.1.2), and compared to six state-of-the-art descriptors for a thorough evaluation. The six descriptors include the Spin Image [14], SHOT [10], RoPS [16], TriSI [29], LFSH [3] and TOLDI [23]. Note that Spin Image is the most cited descriptor in the area of local shape description, and LFSH is the most similar to ours. The SHOT, RoPS and TriSI are verified with an advanced performance [11, 23]. The TOLDI is relatively new proposed descriptor. The parameter settings of the seven feature descriptors are shown in TABLE 1. All these parameter settings, unless otherwise specified, are used for all the experiments in this paper.

To verify the performance of the proposed LPA, we replace the normal used in SI, SHOT and LFSH with our LPA, and then generate three modified descriptors: SI combined with the proposed LPA (SI-LPA), SHOT combined with the proposed LPA (SHOT-LPA), and LFSH combined with the proposed LPA (LFSH-LPA). Note that the parameter settings of SI-LPA, SHOT-LPA and LFSH-LPA are the same with their original descriptors. In addition, for satisfying some time-crucial applications (e.g. robotics, mobile phones, etc.), the comparison of these descriptors in terms of computing efficiency is also tested.

| Parameter Settings for Six Descriptors, where mr Denotes Mesh Resolution. |
|--------------------------------------------------|
| Support radius (mr) | Dimensionality | Length |
| SI | 20 | 15×15 | 64 |
| SHOT | 20 | 8×2×2×11 | 352 |
| RoPS | 20 | 3×3×3×5 | 135 |
| TriSI | 20 | 15×15×3 | 675 |
| LFSH | 20 | 15×10×5 | 30 |
| TOLDI | 20 | 3×20×20 | 1200 |
| SDASS | 20 | 15×5×30 | 345 |

4.3.1 Performance on the BR Dataset

The purpose of testing on the BR dataset is mainly to verify the robustness to noise and varying mesh resolutions. To compactly present the performance, we only use the AUCpr as an evaluation criterion in this section. For each descriptor, we follow the steps introduced in Sect. 4.1.2 to generate AUCpr on the BR dataset. The results of all descriptors are presented in Fig.9.

In the view of robustness to noise (Fig.9 (a)), we give a number of observations. First, the proposed SDASS descriptor outperforms the others in terms of all levels of noise by a large margin and the advantage of our descriptor is more obvious with the increase of noise levels. Second, the descriptor SHOT-LPA achieves the second best performance for resisting noise, and SI and LFSH have the worst performance for resisting noise. Third, as the noise levels increased, the performance of SI, LFSH and SHOT deteriorated sharply. The descriptors SI-LPA, LFSH-LAP and SHOT-LPA correspondingly outperform their original descriptors SI, LFSH and SHOT by a large margin. The main reason for this phenomenon is that the robustness of our LAP to noise significantly outperforms that of the normals (TN and RN) (as verified in Sect.4.2.1).

In terms of robustness to varying mesh resolutions, as shown in Fig.9 (b), several observations can be made from the results. First, our proposed SDASS descriptor outperforms all the other descriptors under all levels of mesh decimation, and the gap is broadened with the increase of mesh decimation. Second, in the range of decimation rate from 10/10 to 4/10, the performance of TriSI descriptor is close to our SDASS descriptor. As the levels of decimation rate surpass 4/10, the performance of TriSI deteriorates sharply and the gap between our SDASS and TriSI is more obvious. Similar to the TriSI, the performance of RoPS and TOLDI also have a significant drop in the high levels of
decimation rate. The common trait of these three descriptors is of using a view-based mechanism. Third, similar to the performance of resisting noise, the descriptors SI-LPA, LFSH-LAP and SHOT-LPA also outperform their original descriptors SI, LFSH and SHOT in terms of resisting varying mesh resolution.

For the robustness to the impact of combining noise and varying mesh resolutions, as shown in Fig.9 (c), our SDASS descriptor also significantly outperform the other methods in all levels of combined noise and mesh decimation (except the highest level). In the highest level, all descriptors are failing to work, as shown in Fig.9 (c). The overall performance of SHOT-LAP is

the second best. The TriSI has a better performance in the low levels of combined noise and mesh decimation, and the superiority is no longer existent in the high levels. The performances of SHOT-LAP, TOLDI, LFSH-LAP, RoPS, SHOT and TriSI are comparable against the performances of SDASS, SI and LFSH descriptors. The performances of SI and LFSH are significantly inferior to others. Their fail to work even under the low level of combined noise and mesh decimation. It is because that the SI and LFSH descriptors use the normal, which has a weak robustness to noise and varying mesh resolutions, as their LRA.

(a) Gaussian noise. (b) Varying mesh resolutions. (c) The combination of Gaussian noise and varying mesh resolution.

Fig.9. AUCpr performance evaluation of the ten feature descriptors tested on the BR dataset. Note that the legends of the subfigure (a) and (b) are same with that of the subfigure (c). The x labels in the subfigure (c) present the Gaussian noises in the subfigure (a) correspondingly combind with the varying mesh resolutions in the subfigure (b).

Fig.10. PRC performance evaluation of the eleven feature descriptors on the four datasets (UWAOR, UWAOR-IR, UWA3M, UWA3M-IR).

4.3.2 Performance on UWAOR, UWAOR-IR, UWA3M and UWA3M-IR Datasets

In this section, all descriptors are tested on UWAOR, UWAOR-IR, UWA3M, UWA3M-IR datasets for verifying their robustness to clutter, occlusion and mesh boundary. The propose of using UWAOR-IR and UWA3M-IR datasets is mainly to present the influence of mesh boundary to the descriptors. The UWAOR-IR and UWA3M-IR datasets are introduced in Sect.
4.1.1. In addition, in consideration of the LRF proposed by Yang et al. [23] has the best performance to resist occlusion, clutter and mesh boundary which are existent in UWAOR and UWA3M datasets (see Sect. 4.2.1), we use the z-axis in this LRF to replace our LRA for generating the SDASS descriptor, and call this new integrated descriptor “SDASS-Yang” in the following.

The performances of all descriptors tested on UWAOR, UWAOR-IR, UWA3M and UWA3M-IR datasets are shown in Fig.10. Several observations can be drawn from these results. First, the descriptors SDASS-Yang and SDASS achieve the best performance on all four datasets, and they outperform others by a large margin. In particular, SDASS-Yang slightly outperform SDASS on UWAOR and UWA3M datasets, and they have a similar performance on UWAOR-IR and UWA3M-IR datasets. It is mainly because that the performance of z-axis in the LRF proposed by Ying et.al [23] outperform our LRA when tested on UWAOR and UWA3M, and they have a similar performance when tested on UWAOR-IR and UWA3M-IR. Second, the performances of all descriptors are universally improved from tested on UWAOR and UWA3M to UWAOR-IR and UWA3M-IR. It implies that mesh boundary has a significant impact to the performance of the descriptors. Third, similar to the results on BR dataset, the performances of SI-LPA, SHOT-LPA and LFSH-LPA are significantly improved by combining our LPA with their original descriptors SI, SHOT and LFSH. It implies that the proposed LPA has a higher robustness to clutter, occlusion and mesh boundary compared to the normals (i.e. TN, RN).

4.3.4 Time Efficiency
In this section, the efficiencies of these descriptors (including SI, SHOT, RoPS, TriSI, LFSH, TOLDI and SDASS) are tested. Since the efficiency is mainly correlate to the number of points in local region, we just use the BR dataset to test these descriptors. First, we randomly select 1000 points in each model of BR dataset (6000 points in total). Second, the total time costs of each descriptor implemented on these points with respect to different support radius R are counted. Here, the values of R are the same with the set in Sect. 4.2.2. The tested results of these descriptors are shown in Fig.11.

![Graph showing computational efficiency of descriptors](image)

Fig.11. The computational efficiency of the seven feature descriptors with respect to varying support radii. The Y axis is shown logarithmically for clarity.

For the results shown in Fig.11, the efficiencies of LFSH, SI, SDASS, TOLDI and SHOT are comparable. Specifically, LFSH and SI rank the first compared to all the other descriptors, and SDASS and TOLDI achieve the secondary calculation efficiency, and SHOT has the worst performance among these five descriptors. TriSI and RoPS have a similar computational performance, and their computational performance is inferior than the others by one order magnitude. In conclusion, our SDASS achieve a better computational performance although it is slightly inferior to the computational performance of LFSH and SI. It is worth noting that RoPS and TriSI descriptors are the two most time-consuming, and their main time consumptions are produced by the time-consuming LRF construction process (as verified in Sect.4.2.1).

5 APPLICATIONS
To further verify the effectiveness of the proposed SDASS descriptor, 3D registration is performed based on the SDASS descriptor. Considering that each local feature descriptor can be applicable to a particular pipeline of 3D registration (e.g. the LRF based descriptors are appropriate to use 1P-RANSAC [41] for searching consistency, while the LRA based descriptors are appropriate to use 2SAC-GC [3]), we next only introduce the application of our SDASS descriptor, and the comparison with other descriptors isn’t involved.

The UWA3M is a popular dataset used for 3D registration [2]. Here, we also simply use the UWA3M dataset for verifying the application of our SDASS descriptor in 3D registration. We first use the SDASS descriptor to perform pairwise registration. Transformation estimation is an important process in pairwise registration, and there are many methods (including RANSAC [42], 1P-RANSAC [41], 2SAC-GC [31]) for estimating transformation. Considering that our SDASS descriptor is generated on LRA, the 2SAC-GC method, which use two feature points and corresponding LRAs for estimating transformation, is appropriate for our descriptor to perform pairwise registration. For specific procedure of performing pair registration, the scene and model involving in a pairwise registration are first uniformly sampled to reduce computational burden. Then, the key points on scene and model are extracted also by uniform sampling. In general, the key points extracted on scene is dense than that extracted on model for guaranteeing the key points on model having corresponding key points on scene. Next, the SDASS descriptor is generated on these key points. For each SDASS feature on model, its closest SDASS feature is searched on scene. The corresponding SDASS features between model and scene are sorted based on the ascending order of their Euclidean distance. The first n correspondences are selected to estimate the transformation using the 2SAC-GC method (n is set to 50 here). An illustration of pairwise registration with respect to each object in the UWA3M dataset is presented in Fig.12. In the figure, the percentage of correct correspondences (PCC) [4] in the 50 selected correspondences are counted. We can observe that our SDASS achieve a very high PCC.

Based on the pairwise registration introduced above, we further present the application of our SDASS descriptor on multi-view registration. We use the method introduced in [2] for performing multi-view registration. The process of multi-view registration mainly divided into three steps: First, feature descriptor is generated on the key points of all partial views; Second, shape is updated by using pairwise registration for gradually adding new partial view; Third, fine registration algorithm [43] and surface reconstruction algorithm [36] are applied to construct a continuous and seamless 3D model. The results of multi-registration for the four objects in UWA3M dataset are presented in Fig.13. It can be seen that our SDASS integrated into the multi-view registration method proposed in [2] achieve a better performance.
6 CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a novel SDASS feature descriptor for 3D local surface description. The prominent advantage of our SDASS descriptor is its high descriptiveness and strong robustness. Moreover, our SDASS has a superior performance in terms of compactness and efficiency. For generating our SDASS descriptor, we also proposed an improved LRA and a robust LPA.

Our LRA is developed from the LRF proposed by Yang et al. [22]. Our LRA achieved a high repeatability in the presence of noise and varying mesh resolutions compared to existing techniques. Our proposed LPA achieved a good performance for resisting various nuisances (e.g., noise, varying mesh resolutions) compared to the two methods of constructing normals (TN and RN). Based on the LPA, the deviation angles between LRA and LPA for encoding the geometrical information of local surface presented a high descriptiveness and strong robustness. Our SDASS descriptor encode the combination of spatial and geometrical information of a local surface. The spatial information is fully encoded by two features (height and projected radial information) on LRA. The geometrical information is encoded by using the deviation angle between LRA and LPA. Owing to the strong robustness and high descriptiveness of LPA, the geometrical information of local surface encoded in our SDASS descriptor presented a superior performance. We performed a set of experiments to assess our SDASS descriptor with respect to a set of different nuisances including noise, varying mesh resolution and mesh boundary, etc. The experimental results showed that our SDASS descriptor outperforms the state-of-the-art methods by a large margin, and obtains high descriptiveness and strong robustness to resisting noise, varying mesh resolution and other variations. At last, our SDASS is applied in 3D registration presented a good performance of its application ability.

There are several interesting research directions for further research. Since local feature description is a fundamental task in 3D computer vision, further exploring the efficient method of constructing superior descriptor is valuable. Considering that the texture of 3D objects is easy obtained by scanner devices (e.g., the Microsoft Kinect device, stereo sensors, etc.), integrating RGB information to the SDASS descriptor would be beneficial when the 3D models exhibit poor geometric features but rich photometric cues. In addition, integrating our SDASS descriptor to specific applications (e.g., 3D object recognition and surface registration) also need to be further researched in the future.

ACKNOWLEDGMENT

The authors would like to acknowledge the Stanford 3D Scanning Repository, the University of Western Australia (UWA) for making their datasets available to us. We also thank the help of Dr. Guo, Dr. Yang and Dr. Tombari for providing their codes to us. This work was partly supported by the National Natural Science Foundation of China (Project No. 51575354), and the Shanghai Municipal Science and Technology project (Project No. 16111106102).

REFERENCES

[1] A. Petrelli and L. Di Stefano, "Pairwise Registration by Local Orientation Cues," Computer Graphics Forum, vol. 35, pp. 59-72, 2016.
[2] Y. Guo, F. Sohel, M. Bennamoun, J. Wan, and M. Lu, "An Accurate and Robust Range Image Registration Algorithm for 3D Object Modeling," IEEE Transactions on Multimedia, vol. 16, pp. 1377 - 1390, 2014.
[3] J. Yang, Z. Cao and Q. Zhang, "A fast and robust local descriptor for 3D point cloud registration," Information Sciences, vol. 346-347, pp. 163-179, 2016.
[4] Y. Guo, B. Mohammed, S. Ferdous, M. Lu, and J. Wan, "3D Object Recognition in Cluttered Scenes with Local Surface Features: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, 2014.
[5] A. Aldoma, F. Tombari, L. D. Stefano, and M. Vincze, "A Global Hypothesis Verification Framework for 3D Object Recognition in Clutter," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, pp. 1383-1396, 2016.
[6] A. S. Mian, M. Bennamoun and R. Owens, "Three-Dimensional Model-Based Object Recognition and Segmentation in Cluttered Scenes," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 1581-1601, 2006.
[7] A. M. Bronstein, M. M. Bronstein, L. J. Guibas, and M. Ovsjanikov, "Shape Google: Geometric Words and Expressions for Invariant Shape Retrieval," ACM Transactions on Graphics, vol. 30, pp. 1-20, 2011.
[8] Y. Guo and Q. Dai, "View-Based 3D Object Retrieval: Challenges and Approaches," IEEE Multimedia, vol. 21, pp. 52-57, 2014.
[9] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, "An efficient 3D face recognition approach using local geometrical signatures," Pattern Recognition, vol. 47, pp. 509-524, 2014.
[10] F. Tombari, S. Salti and L. Di Stefano, "Unique Signatures of Histograms for Local Surface Description," in European Conference on Computer Vision, New York, 2010, pp. 356-369.
[11] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, and N. M. Kwok, "A Comprehensive Performance Evaluation of 3D Local Feature Descriptors," International Journal of Computer Vision, vol. 116, pp. 66-89, 2016.
[12] J. Yang, Q. Zhang and Z. Cao, "The effect of spatial information characterization on 3D local feature descriptors: A quantitative evaluation," Pattern Recognition, vol. 66, pp. 375-391, 2017.
[13] A. E. Johnson, "Spin-Images: A Representation for 3-D Surface
Matching."; Carnegie Mellon University, 1997.

[14] A. E. Johnson and M. Hebert, "Using Spin Images for Efficient Object Recognition in Cluttered 3-D Scenes," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, 1999.

[15] S. Salti, F. Tombari and L. Di Stefano, "SHOT: Unique signatures of histograms for surface and texture description," Computer Vision and Image Understanding, vol. 125, pp. 251-264, 2014.

[16] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan, "Rotational Projection Statistics for 3D Local Surface Description and Object Recognition," International Journal of Computer Vision, vol. 105, pp. 63-86, 2013.

[17] A. Flint, A. Dick and A. van den Hengel, "Thrift: Local 3D Structure Recognition," in 9th Conference on Digital Image Computing Techniques and Applications, 2007, pp. 182-188.

[18] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, "Aligning Point Cloud Views using Persistent Feature Histograms," in IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008, pp. 3384-3391.

[19] A. Petrelli and L. Di Stefano, "On the repeatability of the local reference frame for partial shape matching," in 2011 International Conference on Computer Vision, 2011, pp. 2244 - 2251.

[20] A. Mian, M. Bennamoun and R. Owens, "On the Repeatability and Quality of Keypoints for Local Feature-based 3D Object Retrieval from Cluttered Scenes," International Journal of Computer Vision, vol. 89, pp. 348-361, 2010.

[21] R. B. Rusu, N. Blodow and M. Beetz, "Fast Point Feature Histograms (FPFH) for 3D Registration," in IEEE International Conference on Robotics and Automation, 2009, pp. 3212-3217.

[22] Y. Guo, F. Sohel, M. Bennamoun, J. Wan, and M. Lu, "A novel local surface feature for 3D object recognition under clutter and occlusion," Information Sciences, vol. 293, pp. 196-213, 2015.

[23] J. Yang, Q. Zhang, Y. Xiao, and Z. Cao, "TOLDI: An effective and robust approach for 3D local shape description," Pattern Recognition, vol. 65, pp. 175-187, 2017.

[24] S. Salti, A. Petrell, F. Tombari, and L. Di Stefano, "On the Affinity between 3D Detectors and Descriptors," in 2012 Second Joint 3DIM/3DPVT Conference: 3D Imaging, Modeling, Processing, Visualization & Transmission, 2012, pp. 424-431.

[25] A. Flint, A. Dick and A. van den Hengel, "Local 3D structure recognition in range images," IET Computer Vision, vol. 2, pp. 208-217, 2008.

[26] F. Tombari, S. Salti and L. Di Stefano, "Unique Shape Context for 3D Data Description," in ACM Workshop on 3D Object Retrieval, 2010, pp. 57-62.

[27] A. Frome, D. Huber, R. Kolluri, T. Bulow, and J. Malik, "Recognizing Objects in Range Data Using Regional Point Descriptors," in Proceedings of the European Conference on Computer Vision, 2004, pp. 224-237.

[28] Y. Guo, M. Bennamoun, F. A. Sohel, J. Wan, and M. Lu, "3D Free Form Object Recognition using Rotational Projection Statistics," in Applications of Computer Vision, 2013, pp. 1-8.

[29] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan, "TriSI: A Distinctive Local Surface Descriptor for 3D Modeling and Object Recognition," in 8th International Conference on Computer Graphics Theory and Applications (GRAPP), 2013, pp. 86-93.

[30] N. J. Mitra and A. Nguyen, "Estimating Surface Normals in Noisy Point Cloud Data," in Proceedings of the nineteenth annual symposium on Computational geometry. ACM, 2003, pp. 322-328.

[31] J. Yang, Q. Zhang and Z. Cao, "Multi-attribute statistics histograms for accurate and robust pairwise registration of range images," Neurocomputing, vol. 251, pp. 54-67, 2017.

[32] A. Zabarencu, E. Boyer and R. Horaud, "Keypoints and Local Descriptors of Scalar Functions on 2D Manifolds," International Journal of Computer Vision, vol. 100, pp. 78-98, 2012.

[33] M. Pauly, "Point Primitives for Interactive Modeling and Processing of 3D Geometry," vol. PhD: University of Kaiserslautern, 2003.

[34] R. B. Rusu, "Semantic 3D object maps for everyday manipulation in human living environments." KI- Künstliche Intelligenz, 2010.

[35] K. Jordan and P. Mordohai, "A Quantitative Evaluation of Surface Normal Estimation in Point Clouds," in IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014, pp. 4220-4226.

[36] B. Curless and M. Levoy, "A Volumetric Method for Building Complex Models from Range Images," in 23rd Annual Conference on Computer Graphics and Interactive Techniques, 1996, pp. 303-312.

[37] F. Tombari, S. Salti and L. Di Stefano, "Performance Evaluation of 3D Keypoint Detectors," International Journal of Computer Vision, vol. 102, pp. 198-220, 2013.

[38] A. S. Mian, M. Bennamoun and R. A. Owens, "A Novel Representation and Feature Matching Algorithm for Automatic Pairwise Registration of Range Images," International Journal of Computer Vision, vol. 66, pp. 19-40, 2006.

[39] P. J. Besl and N. D. McKay, "A method for registration of 3-D Shapes," IEEE Transaction on pattern analysis and machine intelligence, vol. 14, pp. 239-256, 1992.

[40] Y. Chen and G. Medioni, "Object Modeling by Registration of Multiple Range Images," Image and vision computing, vol. 10, pp. 145-155, 1992.

[41] Y. Guo, B. Mohammed, S. Ferdous, M. Lu, and J. Wan, "An Integrated Framework for 3D Modeling, Object Detection, and Pose Estimation From Point-Clouds," IEEE Transactions on Instrumentation and Measurement, vol. 60, 2015.

[42] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography." Communications of the ACM, vol. 24, pp. 381-395, 1981.

[43] J. Williams and M. Bennamoun, "Simultaneous Registration of Multiple Corresponding Point Sets," Computer Vision and Image Understanding, vol. 81, pp. 117-142, 2001.