IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 13, 2019, accepted June 24, 2019, date of publication June 27, 2019, date of current version July 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2925364

3D Surface Matching by a Voxel-Based Buffer-Weighted Binary Descriptor

RUQIN ZHOU[1], XIXING LI[2], AND WANSHOU JIANG[1,3]
[1]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China
[2]China National Digital Switching System Engineering and Technological Research Center, Zhengzhou 450000, China
[3]Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430072, China

Corresponding author: Wanshou Jiang (jws@whu.edu.cn)

This work was supported by the National Key R&D Program of China under Grant 2018YFB0504801.

ABSTRACT 3D surface matching by local feature descriptors is a fundamental task in 3D object registration, recognition, and retrieval. In view of 2D projected descriptors' information compression and 3D directly voxelized descriptors' sensitivity to density variation and boundary, this paper first proposes a voxel-based buffer-weighted binary descriptor, named VBBD. After local surfaces of detected keypoints are voxelized, each voxel's buffer region is established, and the buffer-weighted Gaussian kernel density is calculated. If the current voxel's buffer-weighted density is larger than its local surface's average density, the voxel is labeled 1, otherwise, it is 0. The proposed descriptor has several merits: 1) direct acquisition of 3D information without projection, the neighbor information is less compressed. 2) voxels are labeled and binarized according to the buffer-weighted density, which improves the robustness to boundary effect, noise and density variation. Based on the VBBD, a global optimal surface matching method based on a Kuhn_Munkres (KM) algorithm is adopted. The strategy has several advantages: 1) by calculating the descriptors of large-scaled surfaces which are down sampled, the number of the point cloud is reduced, and meanwhile, large-scaled information is obtained; and 2) KM algorithm is adopted to find matching pairs to achieve final maximum weight sum of all matching pairs, which can efficiently avoid local optimum. The experimental results show that, compared with other state-of-the-art descriptors, the VBBD has better descriptiveness and robustness, and the surface matching strategy by the VBBD can achieve both high recall and precision.

INDEX TERMS Voxel, Gaussian kernel density, binary descriptor, Kuhn_Munkres matching.

I. INTRODUCTION

In recent years, artificial intelligence (AI) technologies, which is mainly centered on robots and pattern recognition, have been greatly studied with the development in the data acquisition technology, computer processing capacity and the accumulation of big data [1]. As one of the fundamental but challenging issue in computer vision, robotic and remote sensing [2], [3], 3D surface matching by local feature descriptors has been widely used in various applications [4], for instance, 3D object registration [5], 3D model retrieval [6], [7], 3D object recognition [8], [9], and 3D model reconstruction [10], [11].

The definition of 3D surface matching by local feature descriptors is to find correspondences through two sets of local feature descriptors [12]. Traditionally, this scheme involves two main steps: descriptor generation and feature matching. To obtain a fine matching result, besides a matching strategy which can efficiently reduce or avoid local optimum, a qualified local feature descriptor with high descriptiveness and robustness is also a necessary requirement. The descriptor should contain as much neighbor information as possible to achieve a high discrimination [13]; moreover, it should be robustness to a set of nuisances, such as density variation, noise, and occlusion [14].

The associate editor coordinating the review of this manuscript and approving it for publication was Huimin Lu.

A. RELATED WORK

In addition to some classical descriptors (such as SI [15], 3DSC [16], FPFH [17], SHOT [18], RoPS [19]), a large number of new studies domestic and overseas have emerged in recent five years to improve descriptors' performance. B-SHOT [20] is the first 3D binary feature descriptor developed on the SHOT descriptor, which was generated by quantizing the real valued SHOT descriptor to a binary vector. Although its matching efficiency and storage

requirement are improved, there is a loss of information through quantization [21]. Similarly, Shen *et al.* [22] proposed a hybrid descriptor named Frame-SHOT, which combined global structural frames with local signatures of histogram (SHOT). It is more descriptive than local descriptors and more robustness than global descriptors [22], but the float descriptor requires for more storage [20]. Yang *et al.* [23] proposed a local feature statistic histogram named LFSH, which coded on the local depth, point density, and angles between the normal. However, the statistical-based descriptor inevitably loses neighbor information in quantification. 3DBS [21] is a binary descriptor which encoded normal differences among nearest neighbors of a keypoint, but it is sensitive to noise [3]. Prakhya *et al.* [24] introduced a low dimensional descriptor named 3DHoPD, which respectively projected local surfaces onto three coordinate axes of LRF and quantitatively coded the projected density into 15 bins. It requires dramatically low computational time, but the information is much lost by projected on one dimension. Yang *et al.* [25] proposed a novel triple orthogonal local depth images named TOLDI. The descriptor was obtained by concatenating three local depth images which were projected on three orthogonal planes of LRF. Dong *et al.* [3] proposed a binary shape context called BSC, which calculated Gaussian kernel density and distance on three orthogonal projection planes, and then the BSC was formed by generating random point pairs. BRoPH [26] projected local surfaces on three coordinate axis planes and calculation of depth images and density image. The above three descriptors are 2D projection-based. Although they can effectively reduce the computational complexity and the length of descriptors, but the information is compressed due to the projection. Quan *et al.* [4] proposed a local voxelized structure called LoVS, which encoded the local surfaces into bit strings without computing complex geometric features, but the directly voxelized descriptor is sensitivity to density variation and boundary. Furthermore, there are series of comprehensive summaries in each sub-step [27]–[31], which can provide a quick overview of current research status for new researchers in this field.

However, in terms of feature matching strategies, the researches are still deficiency compared to local descriptors. Generally, there are three essential strategies among current feature matching methods [12]: (1) threshold matching, simply choosing pairs with similarities larger than a threshold; (2) nearest neighbor (NN)-based matching, which directly chooses pairs with the maximum similarity; (3) nearest neighbor distance ratio (NNDR)-based strategies, where the ratio of the maximum similarity to the sub-maximal similarity is required. The research of Mikolajczyk and Schmid [32] showed that the matching strategy of NN and NNDR can achieve better results than the threshold method. In order to improve the accuracy of feature matching, there are some improvements of these essential strategies. For example, Xiong and Han [12] adopted a bidirectional nearest neighbor distance ratio (BNNDR)-based strategy, which can obtain better feature matching than the NN-based and NNDR-based strategy. However, the above strategies do not take the global information into account, and it is easy to trap in local optimum, where many-to-one matching often occurs.

B. OUR CONTRIBUTION

Thus, this paper proposes a global optimum surface matching method by a voxel-based buffer-weighted binary descriptor (VBBD). The main contribution of the paper contains two aspects:

(1) A voxel-based buffer-weighted binary descriptor named VBBD is proposed. Firstly, each local surface of detected keypoints is voxelized; then, the buffer region is established according to each voxel's center, and the buffer-weighted Gaussian kernel density of each voxel is calculated. If the current voxel's buffer-weighted density is larger than the local surface's average density, the voxel is labelled 1, otherwise it is 0. The proposed descriptor has several merits: (a) direct acquisition of 3D information without projection, the third-dimensional information is less compressed. b) buffer region is established and voxels are labelled and binarized according to the buffer-weighted density, which improves the robustness to boundary effect, noise and density variation.

(2) Based on the VBBD, a global optimum feature matching method based on Kuhn_Munkres (KM) algorithm is adopted with a hamming-based distance. The strategy has several advantages: (a) by calculating the descriptors of large-scale surfaces which are sampled, the number of point cloud is reduced, and meanwhile, large-scaled information is obtained; (b) the KM algorithm is adopted to find matching pairs to achieve final maximum weight sum of all matching pairs, which can efficiently avoid local optimum and obtain both high recall and precision.

C. PAPER ORGANIZATION

The paper is organized as follows. Section.II presents the details of VBBD generation and the KM-based feature matching method. In Section.III, the datasets and criterion for accuracy analysis are introduced. The related parameters in the proposed descriptor are analyzed and recommended values are given in Section.IV. The experimental results and evaluations of the proposed binary descriptor and feature matching method are shown in Sections.V. Section.VI shows the performance on registration with RANSAC algorithm. And the effect of different keypoint detectors is analyzed in Section.VII, respectively. Finally, a conclusion of the paper and some future researches are presented in Section.VIII.

II. METHODOLOGY

As mentioned above, the proposed surface matching method by the local descriptor VBBD are composed of two steps:

descriptor generation and feature matching. In this section, the generation of the voxel-based buffer-weighted binary descriptor VBBD (as shown in Figure 1) is introduced in Section A; and then, a KM-based global optimal feature matching method (as shown in Figure 2) is presented in Section B.

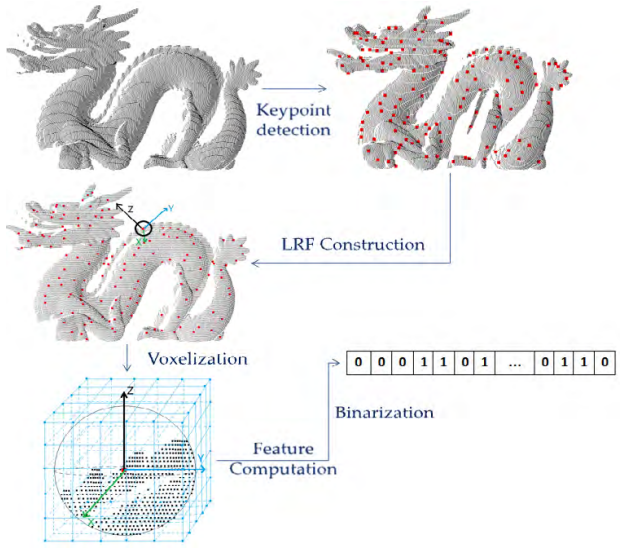


FIGURE 1. Descriptor generation.

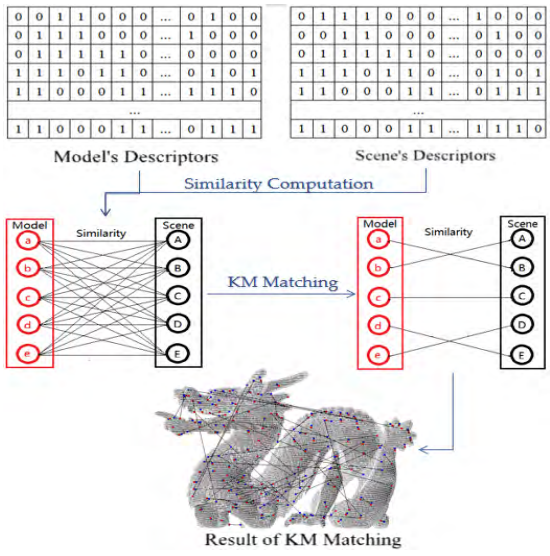


FIGURE 2. Feature matching.

A. VBBD GENERATION

In this section, a voxel-based buffer-weighted binary descriptor named VBBD is introduced. As shown in Fig.1, the scheme of descriptor generation mainly includes five steps: keypoint detection, local reference frame (LRF) construction, voxelization, feature computation and binarization, which are respectively presented in the following sections.

1) KEYPOINT DETECTION

Keypoint detection is a basic but important step in many applications [33], such as registration [5], recognition [10], retrieval [34] and simultaneous location and mapping (SLAM) [35]. An appropriate keypoint detector has a reasonable effect on the descriptors' performance [36] (it will be discussed in the Section.VII). There are two necessary requirements for the detected keypoints: (1) they should be highly repeatable; (2) they should be discriminative and less computational. At present, there are some widely used keypoint detectors, such as voxel sampling, uniform sampling, Harris3D [37], SIFT 3D [38], NARF [39], and ISS [40]. Since the extraction of keypoints is not the focus of this paper, their principles are not described in detail. From experiments in the Section.VII, it is observed that ISS outperforms other keypoint detectors in the Recall versus Precision curve (RP curve). Therefore, through comprehensive consideration [13], ISS is chosen as the keypoint detector for further experiments in this work.

2) LRF CONSTRUCTION

A unique and stable LRF plays a significant role in both robustness and descriptiveness of a descriptor [3]. The invariance to rigid transformation of a descriptor can be achieved by a robust LRF [3], [20]. In this paper, similar to SHOT [18] and B-SHOT [20], the LRF of VBBD is estimated from the eigenvectors of the modified covariance matrix, which is calculated from points in a local surface.

After keypoints p_i are detected, each local surface $S_i = \{q_{ij}, \, || \, q_{ij} - p_i|| < R\}$ of the keypoint p_i is determined by the support radius R to construct the LRF (as shown in Eq.1), where q_{ij} is a point in the local surface of keypoint p_i and$|| \, q_{ij} - p_i||$ is the distance between q_{ij} and p_i.

$$C = \frac{1}{\sum\limits_{q_{ij} \in S_i} (R - ||q_{ij} - p_i||)} \sum_{q_{ij} \in S_i} (R - ||q_{ij} - p_i||)(q_{ij} - p_i)(q_{ij} - p_i)^T \quad (1)$$

Usually, the maximum vector of the eigenvectors is directly chosen as the Z axis, while the minimum is chosen as the X axis. To remove the ambiguity of LRF, the X and Z axes are oriented towards the majority direction of the vectors [25]. And then, the local Y axis is obtained by the cross-product operation of the Z and X axes ($Y = Z \times X$).

3) VOXELIZATION

Voxelization is a good surface representation method with high time and space efficiency [41]. Therefore, voxelization has been widely applied in the robotics and computer vision field [41], [42]. However, in the field of 3D local surface description, limited researches have been conducted [4], as most existing feature descriptors projected 3D local surfaces onto 2D planes for feature coding (such as TOLDI [25], BSC [3], BRoPH [26]) or utilized a spherical volume (such

as SHOT [18], B-SHOT [20] and 3DSC [16]). Quan *et al.* [4] proposed a local voxelized structure (LoVS) for 3D binary feature representation, which has been proved that can effectively handle the challenges existed in low-cost sensors, such as noise and varying data resolutions. Similarly, in the VBBD, voxelization is applied to splitting local surfaces uniformly and efficiently.

After the LRF of each keypoint is constructed, for each keypoint p_i, its local surface S_i are firstly transformed to a new coordinate system based on its LRF [24]. And then, as shown Fig.3, the keypoint p_i is taken as the center, and the transformed local surface $S_i^{'}$ is divided into voxels according to the voxel number ($g * g * g$).

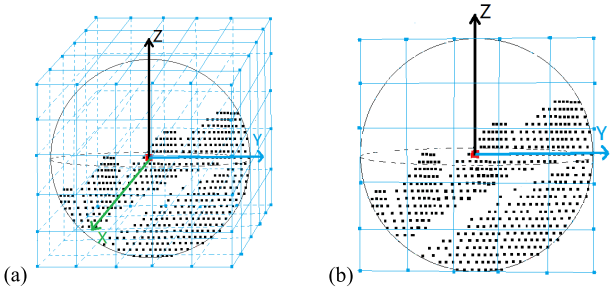


FIGURE 3. Voxelization. (a) is the 3D view. (b) is the front view.

After voxelization, according to a voxel's index (i, j,k), it center's coordinates (x, y, z) can be quickly calculated according to Eq.2, which will be used in further buffer-weighted feature computation.

$$\begin{cases} \text{x = current\_key.x + (i-g / 2) * \_s;} \\ \text{y = current\_key.y + (j- g / 2) * \_s;} \\ \text{z = current\_key.z + (k- g / 2) * \_s;} \end{cases} \tag{2}$$

where *current_key* is the transformed coordinate of the keypoint p_i in current local surface, and the size $\_s$ of each voxel can be obtained as $2 * R/g$.

4) FEATURE COMPUTATION AND BINARYZATION

The commonly used coding strategy for voxels is to directly label 1 or 0 according to whether it contains points or not [4], [41]. However, it is sensitive to density variation and boundary effect. To improve the robustness of the voxel-based descriptor, a Gaussian kernel density, which has been proved to improve the robustness to noise, varying point density, and boundary effect in BSC [3], is adopted for feature calculation in buffer region. It is worth noting that, differed from BSC, VBBD calculates the weighted density of each voxel rather than the weighted density of each point, which is more time efficiency.

Firstly, each voxel's buffer region is established, where the voxel's center is as the origin and the bandwidth h is as the radius (as shown in the yellow circle in Fig.4). Then, the Gaussian kernel density $\psi(i, j,k)$ of current voxel is calculated by points in the buffer region according to Eq.3. Meanwhile, the average Gaussian weighted density ψ_{ave} of all voxels in the local surface $S_i^{'}$ is calculated according to Eq.4. If the current voxel's buffer-weighted density $\psi(i, j,k)$ is larger than the local surface's average density ψ_{ave}, the voxel is labelled 1, otherwise it is 0.

$$\varphi(i, j, k) = \frac{1}{m_i} \sum_{n=1}^{m_i} \frac{1}{\sqrt{2\pi}h} e^{-\frac{||t_n - v_{i,j,k}||^2}{2h^2}}, \quad s.t. ||t_n - v_{i,j,k}|| < h \tag{3}$$

$$\varphi_{\text{ave}} = \frac{\sum_{v_{i,j,k} \in S_i^{'}} \varphi(i, j, k)}{g*g*g} \tag{4}$$

where m_i is the number of points in current buffer region, $v_{i,j,k}$ is the center of voxel (i, j, k), t_n is the neighbor point in the buffer region, $g*g*g$ is the voxel number.

Taking a 2D local surface as an example. Fig.4 illustrates the differences between LoVS and VBBD for the same local surfaces in three different situations: the original local surface (Fig.a), the same surface with boundary effects (Fig.b) and the same surface with density variation (Fig.c). The one differing from the original description is colored in green. It can be seen from the descriptions that in the three situations, there are slight changes (4/25) in LoVS when boundary affected and density varied. However, for VBBD, the descriptions are mostly not change, which shows its robustness to density variation and boundary effect owing to the buffer-weighted feature computation and binarization.

B. KM-BASED FEATURE MATCHING

The task of feature matching is to find a set of correspondences between two sets of feature descriptors [12]. To improve the matching accuracy, descriptors of large-scaled local surfaces are generated after down sampling, therefore, the number of point cloud is reduced, and meanwhile, large-scaled information is obtained. Then, similarities of two sets of descriptors are evaluated by a hamming distance, and finally, the KM algorithm is adopted to find matching pairs to achieve final maximum weight sum of all matching pairs.

1) SIMILARITY CALCULATION

There are many similarity measurements between two descriptors in both mathematics and information science [43]–[45], for example, cosine similarity, Euclidean distance, Dice distance, Jaro-Winkler Distance and Levenshtein distance [45]. A comprehensive survey has been given in [44]. In this paper, a hamming distance is used to measure the similarity, which is time efficiency and easy to implement. For two binary descriptors x and y, the hamming distance is the number of '1' in x XOR y. However, the hamming distance is not proportional to the similarity. To make it proportional, we define the length of the descriptor minus hamming distance as the similarity score.

2) KUHN_MUNKRES ALGORITHM

Feature matching between two sets of descriptors can abstracted as an issue of graph matching [7], which is a basic

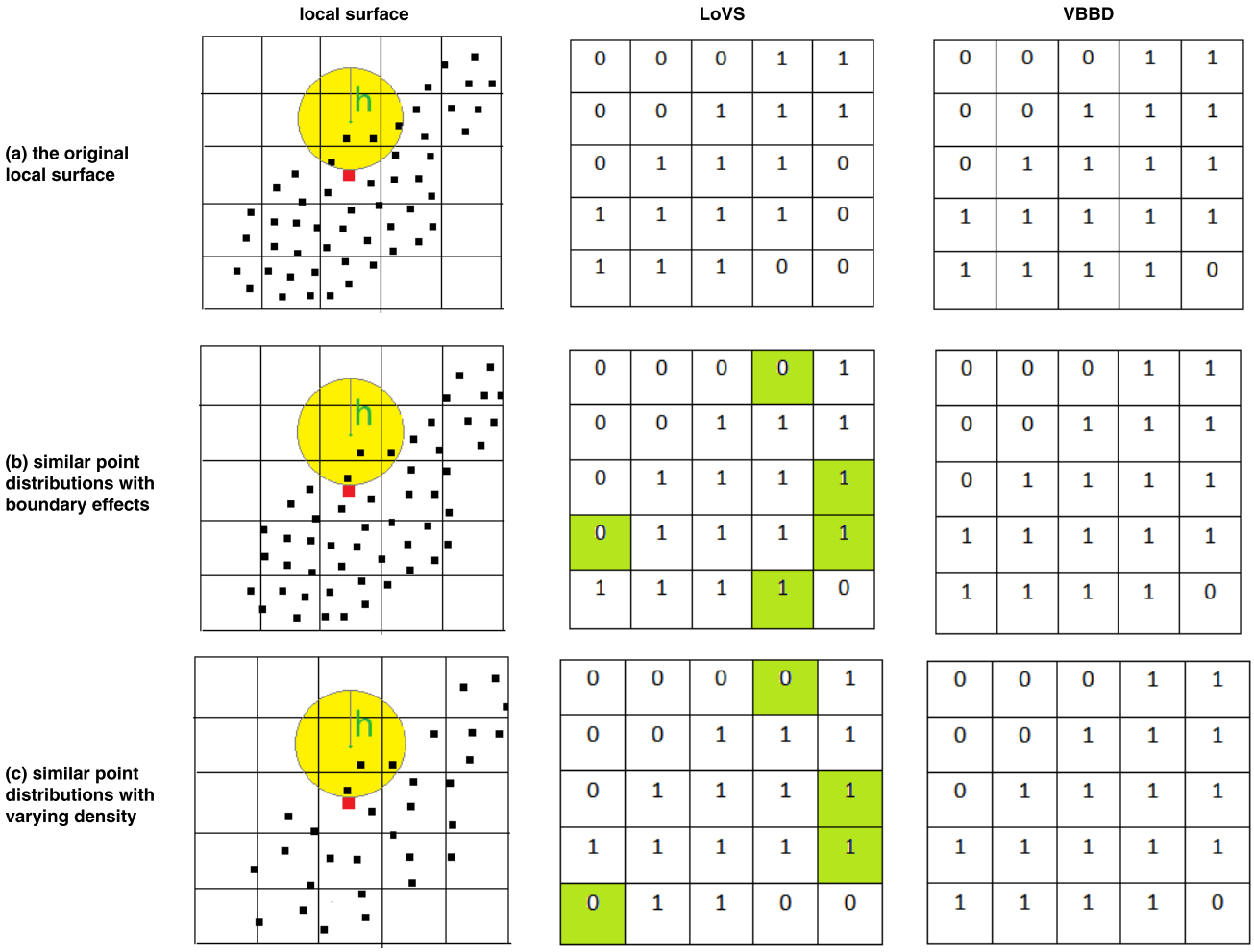


FIGURE 4. Comparison between LoVS and VBBD.

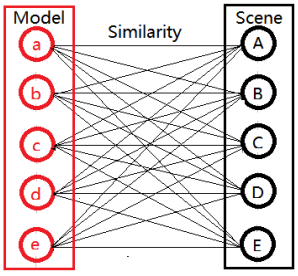


FIGURE 5. Initial matching.

theory in computer field [46]. In this section, a KM algorithm is adopted to find correspondences between two sets of descriptors. As a classical global optimal algorithm to solve the matching problem of weighted bipartite graphs [46], [47], KM algorithm achieves final maximum weight sum of all matching pairs, which can reduce local optimum.

Theoretically, a weighted bipartite graph G is composed of three parts: vertices X, vertices Y, and weighted matrix W. As shown in Fig.5, in this paper, the weight matrix W is the similarities between two sets of descriptors, vertices X are the indexes of model's keypoints, while vertices Y are the indexes of scene's keypoints. It is worth noting that the KM algorithm is suitable for matching between graphs with the same number of nodes. However, the number of keypoints are usually not equal in the model and the scene. In order to adapt it to different node graphs, the low-dimensional data is expanded by adding some vertices or edges of 0 [46], which makes dimensions of both the row and column dimension be the same, and achieves one-to-one matching of all nodes in the algorithm.

Let the labels of vertex $x_i \in X$ and $y_j \in Y$ respectively be A[i] and B[i], and the weight between vertex x_i and y_j be w[i, j]. Initially, for $x_i \in X$, let A[i] is the maximum weight of all weights associated with vertex x_i, while for $y_j \in Y$, B[j] = 0. And then, the schedule of KM algorithm works as follows (Algorithm 1):

III. DATASET AND CRITERIAL

In this section, datasets (including Stanford 3D Scanning Repository and Kinect Views) and criterions (Recall versus Precision curve) to test and assess the proposed feature

Algorithm 1 KM Algorithm

Input: bipartite graph $G=(X,Y,W)$
Output: maximum weighted matching M
(1) Initialize the labels A[i] and B[j] of vertices X and vertices Y;
(2) Adopt Hungarian algorithm [48] to find a perfect matching (as Fig.6);
(3) If perfect matching is found, stop; otherwise, modify the labels' value of vertices;
(4) Repeat (2) (3) until a complete match of equal subgraphs is found.

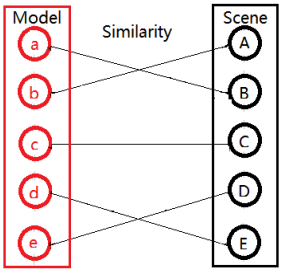


FIGURE 6. Perfect matching.

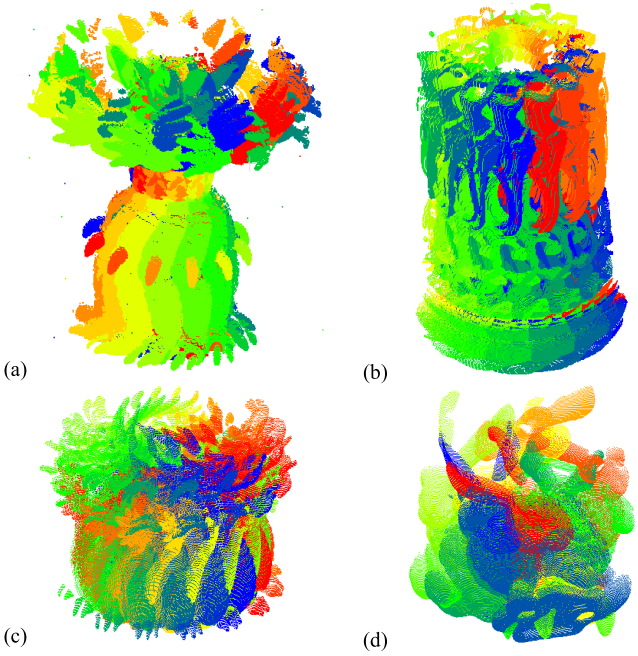


FIGURE 7. Scans of four models in Stanford 3D Scanning Repository. (a) Armadillo; (b) Happy Buddha; (c) Dragon; (d) Bunny.

matching method are respectively introduced in Section.A and Section.B.

A. DATASET

To prove the feasibility of the proposed surface matching method by VBBD, two different kinds of datasets (an overview of benchmark datasets in computer vision has been given in [49]) are utilized: Stanford 3D Scanning Repository (http://graphics.stanford.edu/data/3Dscanrep/) and Kinect Views (http://vision.deis.unibo.it/research/78-cvlab/80-shot).

The Stanford 3D Scanning Repository used in the experiments contains 4 models: Armadillo, Happy Buddha, Bunny and Dragon, which were scanned by a Cyberware 3030 Scanner. For each model, there are several views, for example, stand, side, up, back, and backdrop (Bunny model only has one view). For a fixed view, a set of scans are collected at a fixed rotation angle. In addition, all models are aligned by a modified ICP (Iterative Closest Point) algorithm, and their alignments are stored in ".conf" files with a translation and a rotation. Examples of four models are shown in Fig.7. For better visualization, different scans are colored in different color. More details about the dataset can be found in http://graphics.stanford.edu/data/3Dscanrep/.

The Bologna Kinect Views, collected by a Microsoft Kinect sensor, includes six groups of partial views captured from the Squirrel, Duck, Frog, Mario, Peter Rabbit, and Doll models. Specifically, there are 15, 16, 20, 13, 16, and 15 scans for these models, respectively. Examples of all scans of six models are shown in Fig.8. It is worth noting that only 3D information is utilized and the RGB information is not used. More details about the dataset can be found in http://vision.deis.unibo.it/research/78-cvlab/80-shot.

B. CRITERION

To quantitatively evaluate the performance of the proposed descriptor and surface matching method, Recall versus Precision curve (RP curve) is adopted in the experiments. As one of the widely used measurements in the literature to assess a descriptor [3], [4], [12], the RP curve is calculated as follows: the precision is calculated as the number of correct descriptor matches with respect to the total number of descriptor matches (as shown in Eq.5), while the recall is calculated as the number of correct descriptor matches with respect to the number of corresponding keypoint pairs (as shown in Eq.6). "#" is represented to the total number.

$$\text{precision} = \frac{\#\text{ correct descriptor matches}}{\#\text{descriptor matches}} \tag{5}$$

$$\text{recall} = \frac{\#\text{correct descriptor matches}}{\#\text{ corresponding keypoint pairs}} \tag{6}$$

When the distance $||q_j - Tp_i||$ between the keypoint p_i in a model and the keypoint q_j in a scene is smaller than a threshold s, the p_i and q_j is considered as a corresponding keypoint pair ($\boldsymbol{T}$ is the truth transformation matrix). A descriptor match is considered as correct when the corresponding descriptors' keypoints are corresponding keypoint pairs [3]. By varying the threshold s, a RP curve can be generated. Ideally, if the area under the RP curve is larger, it means that this method achieves both higher recall and precision [29].

FIGURE 8. Bologna kinect views. (a) Squirrel; (b) Duck; (c) Frog; (d) Mario; (e) PeterRabbit; (f) Doll.

IV. PARAMETER ANALYSIS

The proposed VBBD mainly has four parameters: the support radius R for local surface searching, the voxel number g for voxelization, the distance d for down sampling, and the bandwidth h for buffer-weighted density computation. Following, the influence of each parameter will be discussed by a variable-controlling approach, and recommended values of parameters will be given according to experiments.

A. THE SUPPORT RADIUS R FOR LOCAL SURFACE SEARCHING

The support radius R is used to search the neighborhood of keypoints and construct their local surface. It is a crucial parameter in the process of feature descriptor generation. Large-scaled surface can obtain more abundant information, but the computational complexity is higher; small-scaled surface is of lower computational complexity, but the obtained information is limited, and it is sensitive to noise and occlusion [29]. To test the influence of different support radius R on the descriptor, the RP curves of varying support radius R ranging from 5 mr to 70 mr are evaluated with an interval $\Delta = 5$ mr (mr is the model resolution).

Fig.9 shows the RP curves of different support radius R when voxel number $g = 9$, sample distance $d = 1$ mr and bandwidth $h = 2 * \mathrm{R/g}$. It can be seen from the figure that when the support radius is small ($R = 5$ mr), its RP curve is much lower than others. This is because limited information is included in the descriptor, resulting in poor descriptiveness and robustness. With the support radius varying from 5 mr to 60 mr, the recall and precision of the RP curves increase basically. When the support radius is larger than 60 mr, the recall and precision of the RP curve are not improved with the radius increasing, as it becomes more sensitive to partial overlap. Therefore, considering the richness of information and the computational efficiency, the support radius $R = 60$ mr is selected to construct local surfaces in the following experiments.

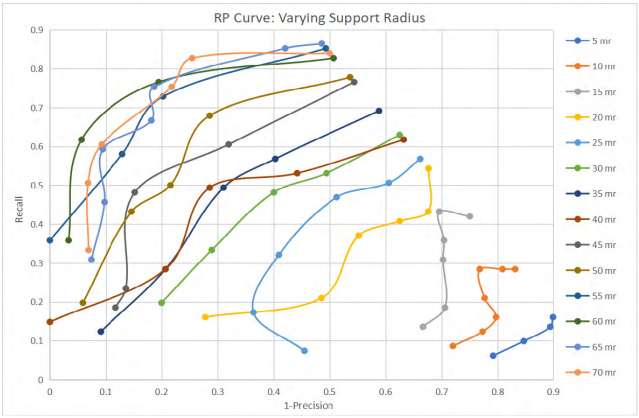


FIGURE 9. RP curve of varying support radius.

B. THE VOXEL NUMBER G FOR VOXELIZATION

The voxel number g for voxelization is directly related to the descriptiveness and robustness of the descriptor. For a given

support radius, if the voxel number is too large, it is of high descriptiveness, but it is not robust enough; when the number for voxelization is small, its robustness is improved, but the descriptiveness is decreased. To find a balance between the descriptiveness and robustness, in this section, varying voxel number g for voxelization are tested from 5 to 17 with an interval $\Delta = 2$.

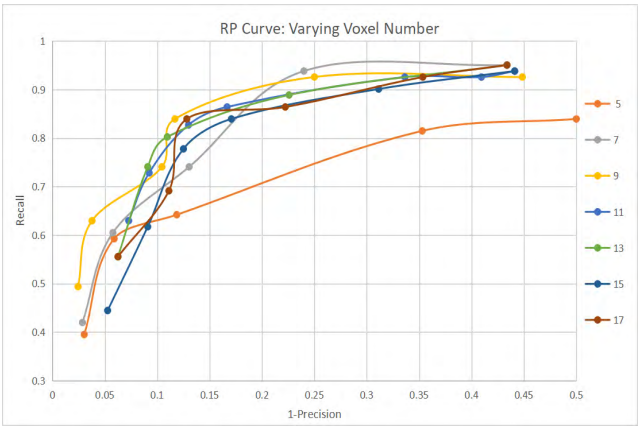


FIGURE 10. RP curve of varying voxel number.

Fig.10 shows the RP curves of varying voxel number when support radius $R = 60$ mr, sample distance $d = 1$ mr and bandwidth $h = 2 * \mathrm{R/g}$. It can be seen that when the voxel number g is 5, its RP curve is lower than other numbers' RP curves. This is because one voxel contains many points, resulting in low descriptiveness. With the voxel number varying from 5 to 9, the recall and precision are increased. This is because more details about the point distribution of the local surface are encoded into the descriptor. However, the performance slightly degrades with the number continues to increase. This is because a large voxel number means a small voxel size, although the local surface feature can be fully expressed, but it is sensitive to noise and voxel partition. On the other way, with voxel number varying from 7 to 17, the differences of RP curves in both recall and precision are less than 0.1, which also shows the method's robustness to voxel partition. Thus, considering both the length of descriptor and the RP curve, we choose g = 9 in the following experiments.

C. THE DISTANCE D FOR DOWNSAMPLING

To obtain more surface information, large-scaled radius is adopted. However, when the radius becomes large, the number of neighbor points increases, which lead to low computational efficiency. Therefore, to reduce the number of point cloud, a uniform sampling of local surfaces is utilized. In order to explore the influence of sampling distance d on the performance of the descriptor, experiments of different sampling distances d varying from 1 mr to10 mr are carried out.

Fig.11 shows the RP curves of different sampling distance d varying from 1 mr to 10 mr, when the support radius

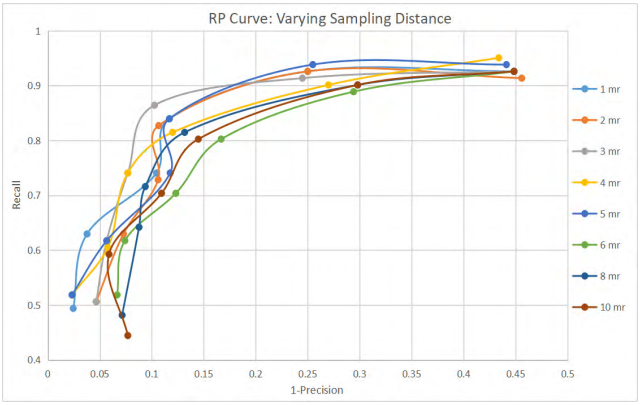


FIGURE 11. RP curve of varying sampling distance.

$R = 60$ mr, the voxel number $g = 9$, and the bandwidth $h = 2 * \mathrm{R/g}$. As can be seen from the graph, the trend of RP curves are basically consistent. With the sampling distance d increasing, the differences among RP curves are slight, which also shows that the descriptor is insensitive to the density variation. However, when the sampling distance d continues to increase, some local surface information is lost, resulting in accuracy decreasing. Therefore, considering both computational efficiency and accuracy of the descriptor, we choose sample distance $d = 5$ mr in the following experiments.

D. THE BANDWIDTH H FOR BUFFER-WEIGHTED DENSITY COMPUTATION

The bandwidth h is mainly used to calculate the Gaussian kernel density in buffer region. Direct voxelization is sensitive to density, boundary effect and voxel partition. By establishing a buffer region, this problem can be effectively reduced. To improve the robustness to Gaussian noise, density variations, boundary effects, and LRF perturbations, the proposed descriptor adopts the Gaussian kernel density for feature computation in buffer region. To test the performance of the varying bandwidth h, experiments are conducted, while the other three parameters are set as $R = 60$ mr and $g = 9$, sampling distance $d = 5$ mr, and $\_size = \mathrm{R/g}$.

Fig. 12 illustrates the RP curves for varying bandwidth h from 1*$\_size$ to 7*$\_size$ with an interval $\Delta = \_size$. As shown in Fig. 13, when $h = 1$*$\_size$, the buffer-weighted density only considers the points inside the current voxel, not considers the information of adjacent points outside the voxel, therefore, the recall and precision are much lower than others. It shows that when the bandwidth h varying from 1*$\_size$ to 4*$\_size$, the larger the buffer region, the better accuracy. And when the buffer region continues to increase from 5*$\_size$ to 7*$\_size$, the performance degrades as too much neighbor information is taken into account. It is obvious that when $h = 4$*$\_size$, the RP curve is obviously higher than other curves. Therefore, in the following experiments, the bandwidth h for buffer-weighted density calculation is set to 4*$\_size$.

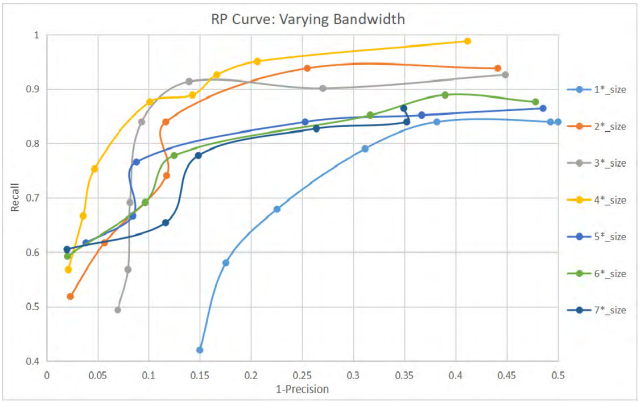


FIGURE 12. RP curve of varying bandwidth.

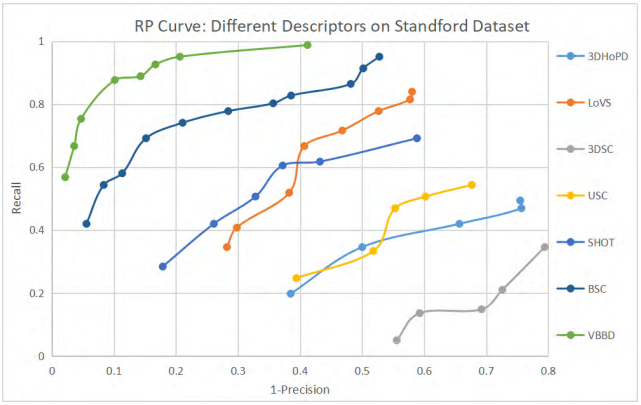


FIGURE 13. RP curves of the stanford 3D scanning repository.

V. PERFORMANCE OF VBBD AND SURFACE MATCHING METHOD

To test the efficiency of the proposed descriptor and feature matching method, experiments are conducted on two datasets, and the descriptiveness and robustness of the descriptor and the global optimum of feature matching strategy are respectively analyzed in Section.A, Section.B and Section C. To provide the fairest comparison as possible, for the compared feature descriptors, any needed parameter (such as support radius, voxel number) is set according to equivalent values of our proposed descriptor analyzed in Section.IV.

A. DESCRIPTIVENESS

A descriptor is of high descriptiveness if it can provide sufficient information to distinguish one local surface from another [3]. In this section, to test the descriptiveness of the proposed VBBD, we compare it against the state-of-the-art descriptors (BSC [3] represents 2D projected descriptors, 3DHoPD [24] represents 1D projected descriptors, LoVS represents 3D descriptors [4], 3DSC [16], USC [50], and SHOT [18] represents classical descriptors) on both the Stanford 3D Scanning Repository and Kinect Views, which has been proved to be efficiently successful in the literature for surface matching. Especially, 3DSC, USC and SHOT have been implemented in Point Cloud Library (PCL, http://docs.pointclouds.org), and 3DHoPD can be download from https://sites.google.com/site/3dhopd/.

Due to the data quality of the two datasets are quite different, the performances of RP curves are different. As shown in Fig.13 and Fig.14, the proposed VBBD descriptor outperforms all the other proposed descriptors by a large margin on both Stanford 3D Scanning Repository and Kinect Views. The differences are caused as follows: (1) the RP curves of 3DSC are much lower than others. This is because only a LRA (Local Reference Axis) is constructed in 3DSC, leading to ambiguity in longitude division direction, while the others all construct a stable LRF, which shows the importance of building a stable LRF for descriptors; (2) the RP curves of 3DHoPD are very low. This is because that

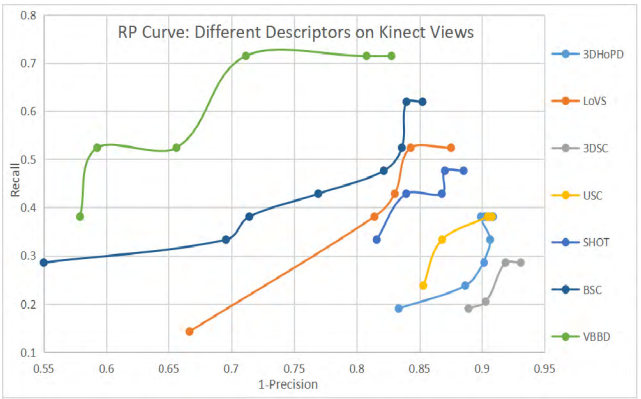


FIGURE 14. RP curves of kinect views.

3D information are greatly compressed by projecting features onto one dimension, while others code features on projected two dimension or direct three dimension. (3) BSC is the 2D projected descriptor, but it outperforms LoVS. This is because the BSC calculates weighted projection density and distance, while LoVS directly labels after voxelization, which is sensitive to density variation and boundary effect. (4) SHOT and VBBD are both three-dimensional descriptors, but the performance are different. This is because SHOT computes the normal' angle, which is sensitive to noise, and its descriptiveness is compressed by quantifying in the process of feature coding. The proposed VBBD extracts 3D information directly, and calculates the Gaussian kernel density in the buffer region, and considers the average weighted density of local surface in the process of binarization, which is more robustness. Therefore, its recall and precision are higher.

B. ROBUSTNESS

A descriptor is of high robustness if it can be insensitive to a number of disturbances that can affect the data, e.g., noise and density variations [36]. To test the robustness of the proposed descriptor, a set of experiments are conducted on Stanford 3D Scanning Repository with respect to a set of disturbances, including different Gaussian noise and varying

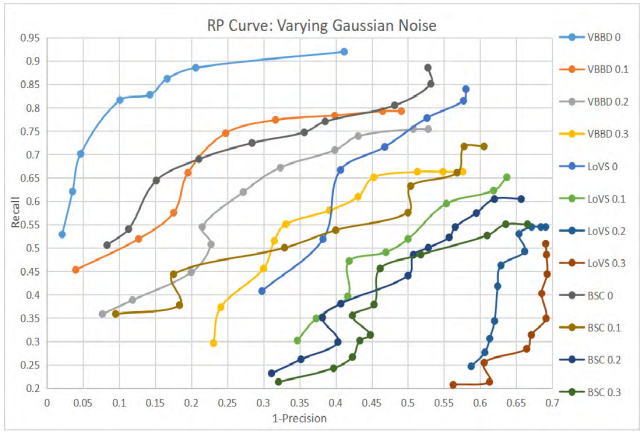


FIGURE 15. RP curve of varying gaussian noise.

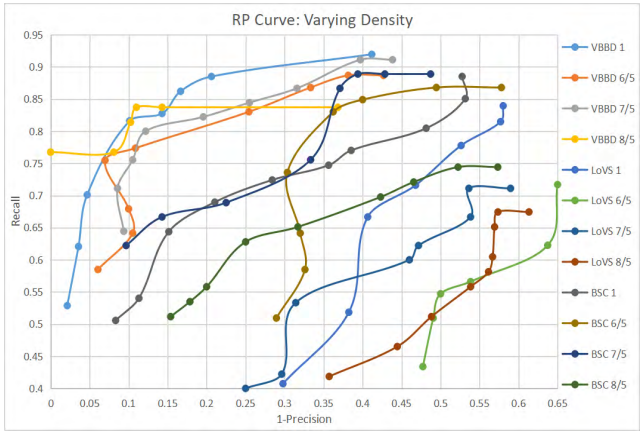


FIGURE 16. RP curve of varying density.

density. The RP curve is also used to evaluate the robustness of the proposed descriptor, with the comparison to the most similar descriptor LoVS [4] and BSC [3].

To evaluate the robustness of the proposed VBBD to different levels of Gaussian noise, five levels of Gaussian noise (0 mr, 0.1 mr, 0.2 mr, 0.3 mr, 0.5 mr) are added on Stanford 3D Scanning Repository. Fig.15 shows the RP curves of an example. As can be seen from the figure, whether it is VBBD, LoVS or BSC, the existence of noise does have a deep impact on the RP curves (on both recall and precision), especially the recall. This is because by adding noise, more points without obvious feature are detected as keypoints, resulting in the increase of "*#corresponding keypoint pairs*" and "*#corresponding matches*", while "*#correct corresponding matches*" is almost unchanged, so the recall and precision of the descriptor are greatly reduced. In addition, the recall decreases with the continuous increase of noise, but the RP curves' trend of the VBBD is consistent, and their decline of recall is less than BSC and LoVS (their recall declines of recall is less than BSC and LoVS (their recall declines greater than 0.3).

To evaluate the robustness of the proposed VBBD to density variation, five levels of density variation (1 mr, 6/5 mr, 7/5 mr, 8/5 mr, 9/5 mr and 2 mr) are sampled on Stanford 3D Scanning Repository. Fig.16 shows the performance of three descriptors, VBBD, LoVS and BSC, at difference density levels. Overall, the trend of all RP curves is consistent as a whole, and the RP curves' fluctuation of BSC and LoVS is obviously larger than VBBD on both recall and precision (> 0.15), which means the density variation has less effect on VBBD than BSC and LoVS. When the precision of VBBD is higher than 0.9, their recall in all RP curves changes a little. When the density varies to 6/5 mr, the RP curves of BSC and LoVS decreases to a certain range. This is because with the density decreases, the local surface becomes flatter, resulting in some details being ignored and their distinctiveness decreasing. However, as the density continues to decrease (7/5 mr), the local surface becomes much smoother, which begins to affect the detection of keypoints. The detected keypoints are

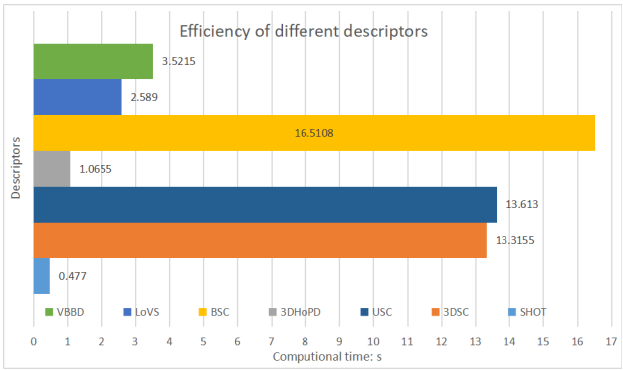


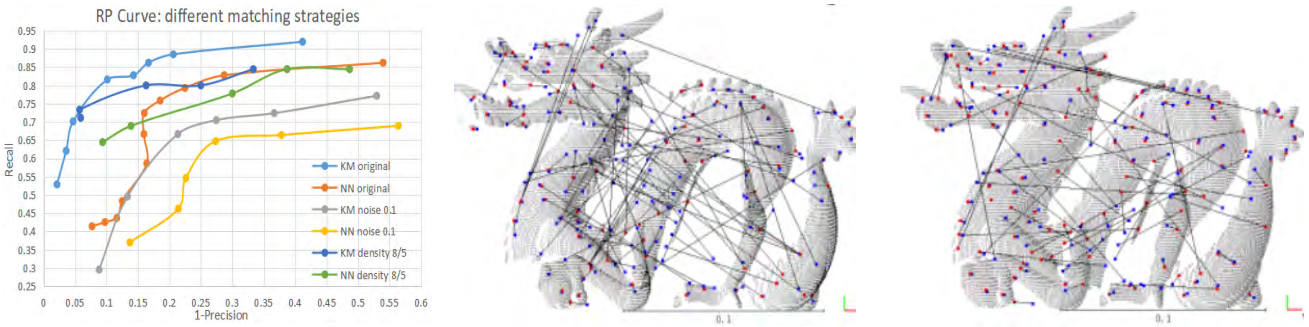
FIGURE 17. Efficiency of different descriptors.

the points with obvious features, and discrimination becomes higher, so the recall slightly increases. The experiments show that the buffer-weighted density computation and binarization improves the descriptor's robustness to density variation.

C. EFFICIENCY

To evaluate the efficiency of the proposed VBBD and surface matching method, experiments are conducted with other descriptors mentioned above. For fairly comparison, the related parameters are set as the same. The detailed information about these descriptors [29] are listed in Tab.1. It is worth noting that the normal used in related descriptors is computed by searching nearest 50 points.

Figure.17 shows the computational time of different descriptors on the same two data, where the number of keypoints on the two data are respectively 163 and 135. As shown in Figure.17, the computational time of BSC, 3DSC and USC are much higher than others. This is because the although BSC is 2D projected descriptor, it computes the weighted density and distance of each points, which need lots of calculations. As for the 3DSC and USC, the bin number is much larger than others, which makes the efficiency decrease. LoVS directly labels 1 or 0 according to whether a voxel contains points or not, without computing



(a) RP Curves. (b) NN matching. (c) KM matching.

FIGURE 18. Performance of different feature matching strategies.

TABLE 1. Information about descriptors.

Descriptor	Dimension	Type	Implementation
SHOT	2*8*2*11 = 352	float	PCL 1.8.0
3DSC	15*12*11 = 1980	float	PCL 1.8.0
USC	15*12*11 = 1980	float	PCL 1.8.0
3HoPD	3*5 = 15	float	C++
BSC	3*2*9*9 = 486	bit	C++
LoVS	9*9*9 = 729	bit	C++
VBBD	9*9*9 = 729	bit	C++

complex feature. Therefore, it is more efficient than other feature-based descriptors. Although both BSC and VBBD compute the weighted feature, the difference is that VBBD calculates the weighted density of each voxel rather than the weighted density of each point, which is more time efficiency than BSC. Additionally, although the support radius is large, through down sampling, the number of point cloud is reduced, and meanwhile, large-scaled information is obtained.

D. GLOBAL OPTIMUM

A feature matching method is considered as global optimum if it comprehensively considers the global information. To test the superiority of the KM matching algorithm compared to NN matching strategy, experiments are conducted on the Stanford 3D Scanning Repository, where the performance of different Gaussian noise and density variation are also considered.

Fig.18 shows the results of KM matching strategy and NN matching strategy with different Gaussian noise and density variation. To fairly compare the differences between the two strategies, the proposed VBBD are generated with the same relevant parameters, and the only difference is the matching strategy. As can be seen from the figures, for the three situations, the recall and precision of KM are both higher than that of NN matching. This is because for NN matching, it only considers the maximum similarity and it is possible to trapped in a local optimum, and it produces a serious many-to-one problem (as shown in Figure a), while for KM matching, it considers the global optimal information. It finds the perfect matching of weighted bipartite graphs, which the final weight sum of all matching pairs can be maximized, and achieves one-to-one matching of all nodes in the algorithm (as shown in Figure b). In addition, when there exists noise interference or density variation, the trend of change coincides with the original two curves.

VI. PERFORMANCE ON REGISTRATION

Registration is one of the important applications for local feature descriptors [51]. The widely used approach for coarse registration is to use corresponding points to estimate the transformation between two point cloud. To verify the validity of the VBBD and KM-based feature matching method, several groups of data on both two datasets were tested with RANSAC registration algorithm, which is implemented in PCL [17] and can obtain both transformation matrix and correspondences after errors eliminated (VBBD for descriptor generation, KM for feature matching, and RANSAC for coarse registration). Theoretically, n models can have $n(n-1)/2$ registration results. Considering the overlap between scans, we choose two continuous scans to register.

Fig.19 and Fig.20 respectively show the examples of registration results on both Stanford 3D Scanning Repository and Kinect Views. For intuitive visualization, the original point cloud is colored in red and the registered point cloud is colored in blue. As can be seen from the graph, after the RANSAC registration, the point cloud of two datasets is basically coincided, even in the case of serious data loss. This is due to the extraction of the large-scale feature and the robustness to data missing. For two scans with larger overlap degree, the registration results are better, which is due to more correspondences with smaller error are matched and they are uniformly distributed. Because the data quality of Kinect Views is not as high as Stanford 3D Scanning Repository (Kinect View's density is smaller, and the local surface is less distinctive), so the registration results are not as good as the Stanford 3D Scanning Repository.

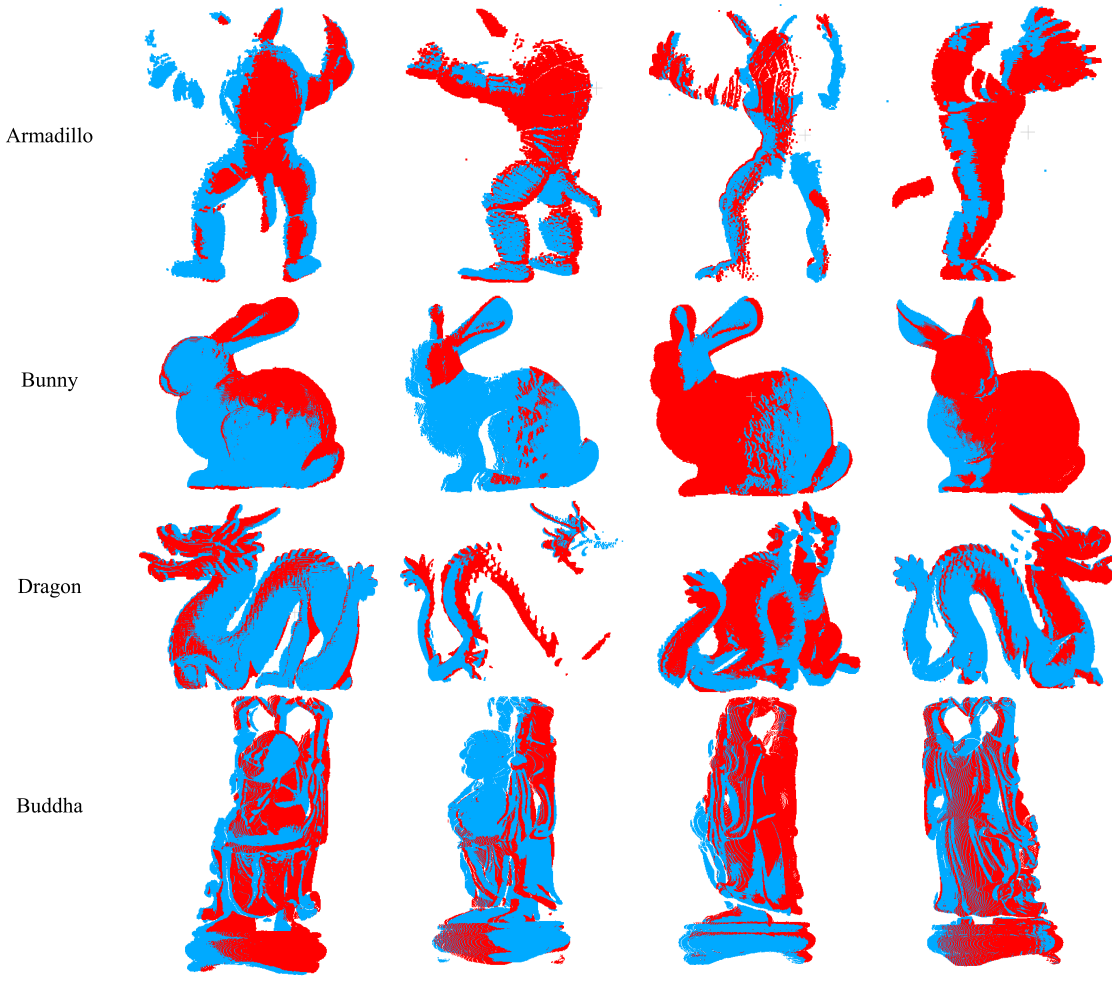


FIGURE 19. Examples of registration results on stanford 3D scanning repository.

For comparison, two methods (VBBD + KM +RANSAC and LoVS + NN + RANSAC) are applied to register all the models in Kinect Views. To quantitatively evaluate the registration result, registration accuracy [14] is introduced for evaluation (as shown in Eq.7). A registration is considered to be correct if the RMSE (Root Mean Squared Error) is smaller than 7 mr. The definition of RMSE is showed in Eq.8. c_s^i and c_t^i are two corresponding points and N is the number of correspondence, R_{GT} and t_{GT} are respectively rotation and translation matrix [4]. "#" is represented to the total number.

$$\text{Reg.accuracy} = \frac{\#\text{correct registration}}{\#\text{registration}} \tag{7}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left\| \text{R}_{\text{GT}} \bullet c_s^i + t_{GT} - c_t^i \right\|^2}{\text{N}}} \tag{8}$$

Tab.2 is the quantitative registration results of Kinect dataset, and Fig.21 is the detailed distribution of the highest recall of all correctly registered scans. As mentioned above, there are 15, 16, 20, 13, 16, and 15 scans for six models on Kinect Views, respectively. Considering the overlap between scans, two continuous scans are chosen to register. Therefore, for each models, the total registration number is 14, 15, 19, 12, 15 and 14. It can be seen from Tab.1 that the method I (VBBD + KM + RANSAC) basically outperforms the method II (LoVS + NN + RANSAC) in the registration accuracy, especially on the doll and mario models. For the accuracy distribution of all correct registrations, as shown in Fig.21, in general, the method I (VBBD + KM + RANSAC) outperforms the method II (LoVS + NN + RANSAC) on mostly situations, as there are 45 bars where yellow is above, and 6 bars where only single yellow. It is worth noting that, in Fig.21, if the yellow is above and the blue is below in a bar, it means the method I outperforms the method II; conversely, it means the method II outperforms the method I. If the bar is only yellow, it means the method I outperforms the method II; if the bar is only blue, it means the method II outperforms the method I. The black bar means the recall of two descriptors is nearly equal.

VII. DISCUSSION

A. EFFECT OF DIFFERENT KEYPOINT DETECTORS

Through experiments, it is found that different keypoint detectors have an important impact on performance of the

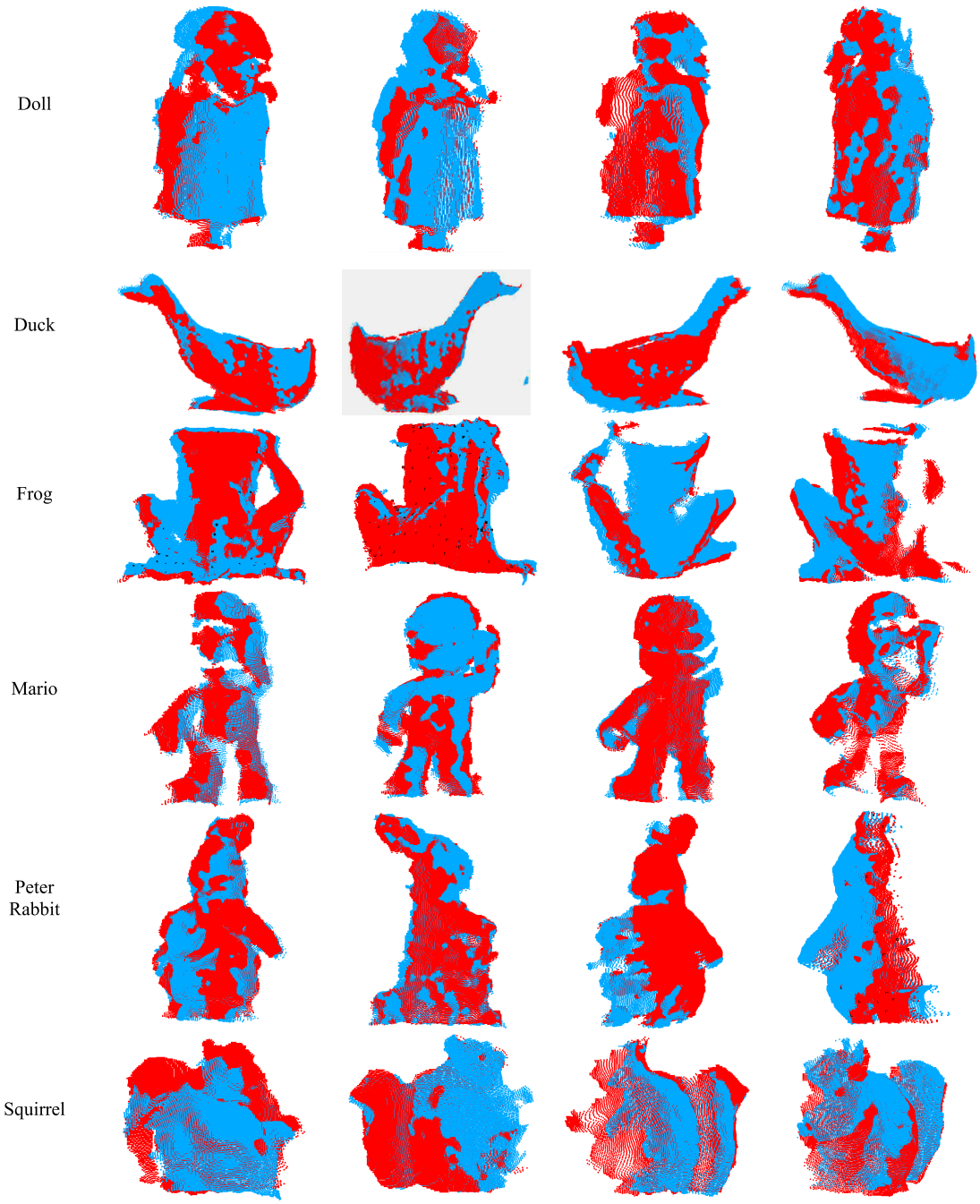


FIGURE 20. Examples of registration results on kinect views.

descriptor and further feature matching. To test their effect on RP curves, several widely used keypoint detectors (ISS, voxel sampling, uniform sampling, SIFT 3D, NARF, Harris3D) mentioned before are chosen as examples. It is worth noting that all the above keypoint detectors are implemented in PCL.

Fig.22 is the results of six keypoint detectors with mostly the same number of keypoints. Experimental results show that the distribution of keypoints detected by ISS, voxel sampling and uniform sampling detectors is more uniform than NARF and Harris3D detectors, especially in surfaces with less features. For SIFT 3D detector, besides the regions with obvious features, the keypoints are mainly concentrated on the edge of the model.

For comprehensive comparison, the time efficiency of different keypoint detector is recorded. As can be seen in Tab.3, to detect the same number of keypoints, the computational time of the six detectors are quite different: the two sampling detectors (voxel sampling and uniform sampling) are the two most efficient detectors, as their principle are simple and the amount of calculation is small. The SIFT 3D is much more

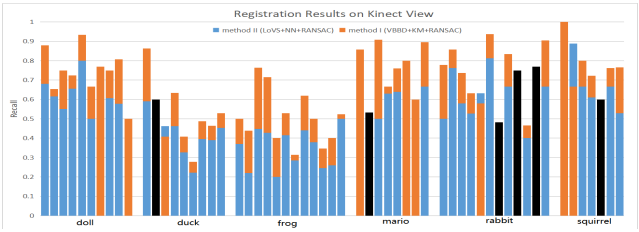


FIGURE 21. Distribution of registration accuracy on kinect views.

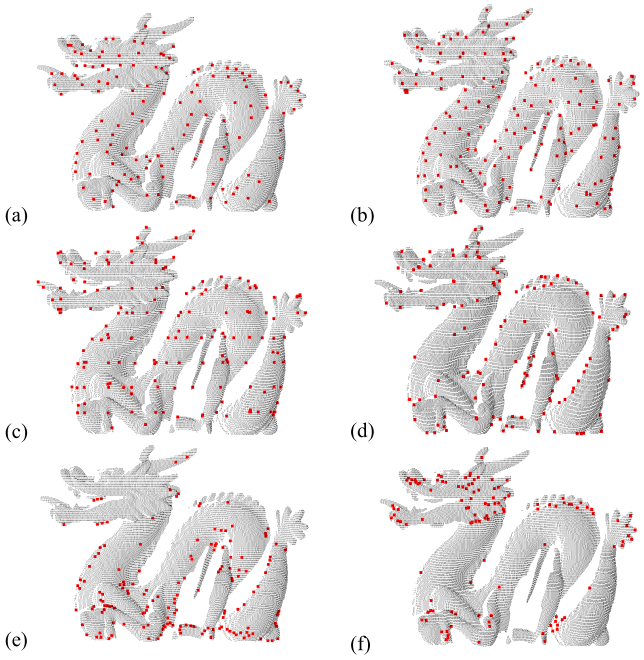


FIGURE 22. Results of different keypoint detectors. (a) is the ISS detector; (b) is the voxel sampling; (c) is the uniform sampling; (d) is the SIFT 3D detector; (e) is the NARF detector; (f) is the Harris3D detector.

complex as the characteristics in multi-scale spaces need to be computed.

Fig.23 shows the analysis of these keypoint detectors on RP curves and the relationships between repeatability, recall and Distance. Fig.a indicates the keypoints' repeatability [33]

TABLE 2. Registration accuracy of kinect view.

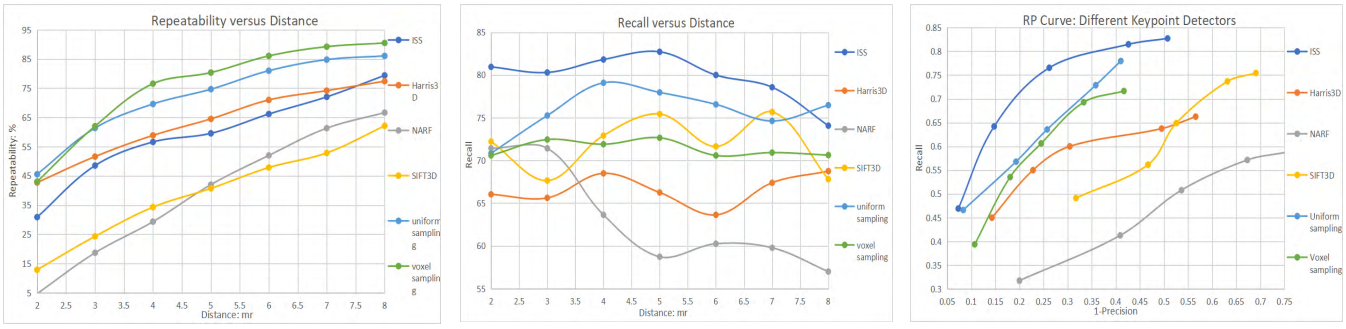
Model	Method II (LoVS +NN+RANSAC)	Method I (VBBD +KM +RANSAC)
doll	0.571429	0.714286
duck	0.666667	0.666667
frog	0.578947	0.578947
mario	0.5	0.666667
rabbit	0.8	0.8
squirrel	0.428571	0.5

TABLE 3. Efficiency of different keypoint detectors.

Detector	Keypoint number	Times (s)
ISS	135	0.3635
Voxel	167	0.002
Uniform	167	0.0035
SIFT 3D	138	0.7295
NARF	159	0.1455
Harris3D	124	0.18

with varied threshold values (distance). Obviously, with the distance increasing, their repeatability is nearly in proportion to the distance. Fig.b shows the relationship between the recall and varying distance of different keypoint detectors. It can be seen that, in different threshold values (distance), their recall fluctuates greatly, except voxel sampling detector. Fig.c shows the effect of different keypoint detectors on RP curves. It is obvious that, although uniform sampling and voxel sampling have high repeatability, their recall and precision are not as high as ISS. This is because many of the keypoints detected by voxel or uniform sampling are not the points with obvious features, which is hard to distinguish. However, their performances are better than SIFT 3D, NARF and Harris3D, which thanks to the uniform distribution of detected keypoints.

Therefore, to obtain a better performance of descriptor generation and feature matching, when detecting keypoints, there are two basic requirements: (1) the detected keypoints should be the points with obvious features as far as possible;



(a) (b) (c)

FIGURE 23. Analysis of different keypoint detectors. (a) repeatability versus distance; (b) recall versus distance; (c) RP curves.

(2) the detected keypoints should be uniformly distributed. For comprehensive consideration, for those objects with obvious feature, such as Happy Buddha, ISS detector is better, while for objects with smooth surface, such as Bunny model and Kinect dataset, uniform sampling is better.

VIII. CONCLUSION

In this paper, a global optimum surface matching method by a voxel-based buffer-weighted binary descriptor (VBBD) is proposed. This method has several advantages: (1) the proposed VBBD is of high descriptiveness, as its direct acquisition of 3D information without projection, the third-dimensional information is less compressed; (2) the proposed VBBD is of high robustness to voxel partition, noise and density variation, as voxels are labelled and binarized according to the buffer-weighted Gaussian kernel density; (3) the KM-based global optimum feature matching method considers the global information and finally achieves a maximum weight sum of all matching pairs, which can efficiently avoid many-to-one matching. However, there are also some shortcomings. One is that there are four parameters in this method: the support radius R for local surface searching, the voxel number g for voxelization, the distance d for down sampling, and the bandwidth h for buffer-weighted density computation. Experiments show that the performance of different values of parameters varies greatly. Inappropriate parameter settings will cause a bad effect for further 3D surface matching. Additionally, the parameter settings are varied for different datasets. Therefore, in our future work, self-adaptive parameter setting method will be studied.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the editors and the reviewers for their constructive and helpful comments for the substantial improvement of this paper.

REFERENCES

[1] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa "Brain intelligence: Go beyond artificial intelligence," *Mobile Netw. Appl.*, vol. 23, no. 2, pp. 368–375, Apr. 2018.

[2] P. Cirujeda, Y. D. Cid, X. Mateo, and X. Binefa, "A 3D scene registration method via covariance descriptors and an evolutionary stable strategy game theory solver," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 306–329, Dec. 2015. doi: 10.1007/s11263-015-0820-2.

[3] Z. Dong, B. Yang, Y. Liu, F. Liang, B. Li, and Y. Zang, "A novel binary shape context for 3D local surface description," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 431–452, Aug. 2017. doi: 10.1016/j.isprsjprs.2017.06.012.

[4] S. Quan, J. Ma, F. Hu, B. Fang, and T. Ma, "Local voxelized structure for 3D binary feature representation and robust registration of point clouds from low-cost sensors," *Inf. Sci.*, vol. 444, pp. 153–171, May 2018. doi: 10.1016/j.ins.2018.02.070.

[5] Z. Dong, B. Yang, F. Liang, R. Huang, and S. Scherer, "Hierarchical registration of unordered TLS point clouds based on binary shape context descriptor," *ISPRS J. Photogramm. Remote Sens.*, vol. 144, pp. 61–79, Oct. 2018. doi: 10.1016/j.isprsjprs.2018.06.018.

[6] S. E. Naffouti, Y. Fougerolle, I. S. A. Aouissaoui, and F. Mériaudeau, "Heuristic optimization-based wave kernel descriptor for deformable 3D shape matching and retrieval," *Signal, Image Video Process.*, vol. 12, no. 5, pp. 915–923, Jul. 2018. doi: 10.1007/s11760-018-1235-7.

[7] H. Tabia and H. Laga, "Covariance-Based Descriptors for Efficient 3D Shape Matching, Retrieval, and Classification," *IEEE Trans. Multimedia*, vol. 17, pp. 1591–1603, 2015. doi: 10.1109/TMM.2015.2457676.

[8] A. Abbad, K. Abbad, and H. Tairi, "3D face recognition: Multi-scale strategy based on geometric and local descriptors," *Comput. Electr. Eng.*, vol. 70, pp. 525–537, Aug. 2018. doi: 10.1016/j.compeleceng.2017.08.017.

[9] B. Yang, Y. Liu, Z. Dong, F. Liang, B. Li, and X. Peng, "3D local feature BKD to extract road information from mobile laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 329–343, Aug. 2017. doi: 10.1016/j.isprsjprs.2017.06.007.

[10] S. A. A. Shah, M. Bennamoun, and F. Boussaid, "Keypoints-based surface representation for 3D modeling and 3D object recognition," *Pattern Recognit.*, vol. 64, pp. 29–38, Apr. 2017. doi: 10.1016/j.patcog.2016.10.028.

[11] Y. Guo, F. Sohel, M. Bennamoun, J. Wan, and M. Lu, "An accurate and robust range image registration algorithm for 3D object modeling," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1377–1390, Aug. 2014. doi: 10.1109/TMM.2014.2316145.

[12] F. Xiong and X. Han, "A 3D surface matching method using keypoint- based covariance matrix descriptors," *IEEE Access*, vol. 5, pp. 14204–14220, 2017. doi: 10.1109/ACCESS.2017.2727066.

[13] Y. Zhong, "Intrinsic shape signatures: A shape descriptor for 3D object recognition," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops, (ICCV)*, Sep./Oct. 2009, pp. 689–696.

[14] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan, "3D object recognition in cluttered scenes with local surface features: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2270–2287, Nov. 2014.

[15] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999.

[16] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2004, pp. 224–237.

[17] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2009, pp. 3212–3217.

[18] F. Tombari, S. Salti, and L. Di stefano, "Unique signatures of histograms for local surface description," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 356–369.

[19] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan, "Rotational projection statistics for 3D local surface description and object recognition," *Int. J. Comput. Vis.*, vol. 105, no. 1, pp. 63–86, Oct. 2013.

[20] S. M. Prakhya, B. Liu, and W. Lin, "B-SHOT: A binary feature descriptor for fast and efficient keypoint matching on 3D point clouds," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep./Oct. 2015, pp. 1929–1934.

[21] S. Srivastava and B. Lall, "3D binary signatures," in *Proc.10th Indian Conf. Comput. Vision, Graph. Image Process.*, Dec. 2016, p. 7. doi: 10.1145/3009977.3010009.

[22] Z. Shen, X. Ma, and X. Zeng, "A Hybrid 3D descriptor with global structural frames and local signatures of histograms," *IEEE Access*, vol. 6, pp. 39261–39272, 2018. doi: 10.1109/ACCESS.2018.2856866.

[23] J. Yang, Z. Cao, and Q. Zhang, "A fast and robust local descriptor for 3D point cloud registration," *Inf. Sci.*, vols. 346–347, pp. 163–179, Jun. 2016. doi: 10.1016/j.ins.2016.01.095.

[24] S. M. Prakhya, J. Lin, V. Chandrasekhar, W. Lin, and B. Liu, "3DHoPD: A fast low-dimensional 3-D descriptor," *IEEE Robot. Autom. Lett.*, vol. 2, no. 3, pp. 1472–1479, Jul. 2017.

[25] J. Yang, Q. Zhang, Y. Xiao, and Z. Cao, "TOLDI: An effective and robust approach for 3D local shape description," *Pattern Recognit.*, vol. 65, pp. 175–187, May 2017. doi: 10.1016/j.patcog.2016.11.019.

[26] Y. Zou, X. Wang, T. Zhang, B. Liang, J. Song, and H. Liu, "BRoPH: An efficient and compact binary descriptor for 3D point clouds," *Pattern Recognit.*, vol. 76, pp. 522–536, Apr. 2018. doi: 10.1016/j.patcog.2017.11.029.

[27] J. Yang, Y. Xiao, and Z. Cao, "Toward the repeatability and robustness of the local reference frame for 3D shape matching: An evaluation," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3766–3781, Aug. 2018. doi: 10.1109/TIP.2018.2827330.

[28] A. Mian, M. Bennamoun, and R. Owens, "On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes," *Int. J. Comput. Vis.*, vol. 89, nos. 2–3, pp. 348–361, Sep. 2010. doi: 10.1007/s11263-009-0296-z.

[29] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, and N. M. Kwok, "A comprehensive performance evaluation of 3D local feature descriptors," *Int. J. Comput. Vis.*, vol. 116, no. 1, pp. 66–89, Jan. 2016. doi: 10.1007/s11263-015-0824-y.
[30] J. Yang, Q. Zhang, and Z. Cao, "The effect of spatial information characterization on 3D local feature descriptors: A quantitative evaluation," *Pattern Recognit.*, vol. 66, pp. 375–391, Jun. 2017.
[31] A. Dittrich, M. Weinmann, and S. Hinz, "Analytical and numerical investigations on the accuracy and robustness of geometric features extracted from 3D point cloud data," *ISPRS J. Photogramm. Remote Sens.*, vol. 126, pp. 195–208, Apr. 2017. doi: 10.1016/j.isprsjprs.2017.02.012.
[32] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*. vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[33] F. Tombari, S. Salti, and L. Di Stefano, "Performance evaluation of 3D keypoint detectors," *Int. J. Comput. Vis.*, vol. 102, nos. 1–3, pp. 198–220, Mar. 2013. doi: 10.1007/s11263-012-0545-4.
[34] N. Markuš, I. Pandžic, and J. Ahlberg, "Learning local descriptors by optimizing the keypoint-correspondence criterion: Applications to face matching, learning from unlabeled videos and 3D-shape retrieval," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 279–290, Jan. 2019. doi: 10.1109/TIP.2018.2867270.
[35] F. Kallasi, D. L. Rizzini, and S. Caselli, "Fast keypoint features from laser scanner for robot localization and mapping," *IEEE Robot. Autom. Lett.*, vol. 1, no. 1, pp. 176–183, Jan. 2016. doi: 10.1109/LRA.2016.2517210.
[36] S. Salti, F. Tombari, and L. Di Stefano, "A performance evaluation of 3D keypoint detectors," In *Proc. Int. Conf. 3D Imag., Modeling, Process., Vis. Transmiss.*, May 2011, pp. 236–243.
[37] I. Sipiran and B. Bustos "Harris 3D: A robust extension of the Harris operator for interest point detection on 3D meshes," *Visual Comput.*, vol. 27, no. 11, pp. 963–976, Nov. 2011.
[38] A. Flint, A. Dick, and A. van den Hengel, "Thrift: Local 3D structure recognition," in *Proc. DICTA*, vol. 7, Dec. 2007, pp. 182–188.
[39] B. Steder, R. B. Rusu, K. Konolige, and W. Burgard, "Point feature extraction on 3D range scans taking into account object boundaries," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2011, pp. 2601–2608.
[40] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE multimedia*, vol. 19, no. 2, pp. 4–10, Feb. 2012.
[41] E. Lemus, E. Bribiesca, and E. Garduño, "Representation of enclosing surfaces from simple voxelized objects by means of a chain code," *Pattern Recognit.*, vol. 47, no. 4, pp. 1721–1730, Apr. 2014. doi: 10.1016/j.patcog.2013.11.002.
[42] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46–57, Jun. 1989.
[43] M. F. Gafar and E. E. Hemayed, "Surface area distribution descriptor for object matching," *J. Adv. Res.*, vol. 1, no. 3, pp. 233–241, Jul. 2010. doi: 10.1016/j.jare.2010.06.005.
[44] S. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *Int. J. Math. Models Methods Appl. Sci.*, vol. 1, no. 4, pp. 300–307, 2007.
[45] Navarro, G. "A guided tour to approximate string matching," *ACM Comput. Surv.*, vol. 33, no. 1, pp. 31–88, Mar. 2001.
[46] Y. Zeng, X. Wu, and J. Cao, "An improved KM algorithm for computing structural index of DAE system," in *Proc. 12th Int. Symp. Distrib. Comput. Appl. Bus., Eng. Sci.*, Sep. 2013, pp. 95–99.
[47] Z. Gao, D. Chen, P. Sun, and S. Cai, "KM-based efficient algorithms for optimal packet scheduling problem in celluar/infostation integrated networks," *Ad Hoc Netw.*, vol. 77, pp. 84–94, Aug. 2018. doi: 10.1016/j.adhoc.2018.05.001.
[48] H. Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logistics Quart.*, vol. 2, nos. 1–2, pp. 83–97, Mar. 1955.
[49] Y. Guo, J. Zhang, M. Lu, J. Wan, and Y. Ma, "Benchmark datasets for 3D computer vision," in *Proc. IEEE 9th Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2014, pp. 1846–1851.
[50] F. Tombari, S. Salti, and L. D. Stefano, "Unique Shape Context for 3D Data Description," in *Proc. ACM Workshop 3D Object Retr.*, Firenze, Italy. vol. 25, Oct. 2010, pp. 57–62.
[51] A. L. Kleppe, L. Tingelstad, and O. Egeland"Coarse alignment for model fitting of point clouds using a curvature-based descriptor," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 2, pp. 811–824, Apr. 2019. doi: 10.1109/TASE.2018.2861618.

RUQIN ZHOU was born in China, in 1995. She received the B.E. degree in remote sensing from Wuhan University, Wuhan, China, in 2017, where she is currently pursuing the Ph.D. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing. Her current research interests include 3D registration, 3D reconstruction, and 3D object recognition.

XIXING LI was born in China, in 1995. He is currently pursuing the M.S. degree from the China National Digital Switching System Engineering and Technological Research Center, China. His current research interests include computer algorithms study and vulnerability mining in software and systems.

WANSHOU JIANG was born in China, in 1967. He received the B.E., M.S., and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1989, 1996, and 2004, respectively, where he is currently a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing. His current research interests include information extraction, image matching and registration, and 3D reconstruction.