IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 22, 2016, accepted January 13, 2017, date of publication February 6, 2017, date of current version March 13, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2658681

3D Object Recognition in Cluttered Scenes With Robust Shape Description and Correspondence Selection

KEKE TANG, (Student Member, IEEE), PENG SONG, AND XIAOPING CHEN
University of Science and Technology of China, Hefei 230027, China

Corresponding author: P. Song (songpeng@ustc.edu.cn)

This work was supported in part by the Anhui Provincial Natural Science Foundation under Grant 1508085QF122, in part by the National Natural Science Foundation of China under Grant 61403357, and in part by the Microsoft Research Asia Collaborative Research Program.

ABSTRACT Recognizing 3-D objects in cluttered scenes is a challenging task. Common approaches find potential feature correspondences between a scene and candidate models by matching sampled local shape descriptors and select a few correspondences with the highest descriptor similarity to identify models that appear in the scene. However, real scans contain various nuisances, such as noise, occlusion, and featureless object regions. This makes selected correspondences have a certain portion of false positives, requiring adopting the time-consuming model verification many times to ensure accurate recognition. This paper proposes a 3-D object recognition approach with three key components. First, we construct a *Signature of Geometric Centroids* descriptor that is descriptive and robust, and apply it to find high-quality potential feature correspondences. Second, we measure geometric compatibility between a pair of potential correspondences based on isometry and three angle-preserving components. Third, we perform effective correspondence selection by using both descriptor similarity and compatibility with an auxiliary set of ''less'' potential correspondences. Experiments on publicly available data sets demonstrate the robustness and/or efficiency of the descriptor, selection approach, and recognition framework. Comparisons with the state-of-the-arts validate the superiority of our recognition approach, especially under challenging scenarios.

INDEX TERMS 3-D object recognition, shape descriptor, correspondence selection, shape matching.

I. INTRODUCTION

With recent advances in 3D geometry acquisition technology, object recognition working with 3D data (e.g., depth scans) receives increasing attention in computer vision and graphics. Compared with traditional 2D object recognition, 3D object recognition is less affected by variation of illumination, shadow, and object texture, while allowing more accurate 6 degree-of-freedom (DOF) pose estimation, thanks to the 3D shape information contained in the data. Nevertheless, 3D object recognition in cluttered scenes, especially under conditions of occlusion and/or noise, remains a challenging research problem [1].

Existing 3D object recognition approaches generally compute local invariant descriptors around sampled feature points on a scene and candidate models, and match scene and model feature points successively via their associated descriptors to attain potential feature correspondences. A few feature correspondences with the highest descriptor similarity are further selected for identifying models that appear in the scene. However, due to nuisances in the scans and/or limited descriptiveness of the descriptors, the selected correspondences are not guaranteed to be correct and usually contain a certain portion of false positives (i.e., incorrect correspondences), see Figure 1 for examples. To ensure accurate recognition, model verification needs to be conducted for each model hypothesis voted by the feature correspondence(s).

To verify a hypothesized model, transforms are generated via the associated corresponding feature point(s) to align the object model to the scene. Then the overlap region between the transformed object model and the scene is estimated. The object is considered to appear in the scene (i.e., be recognized) if the overlap size is larger than a threshold. However, computing overlap between two 3D shapes is computational expensive, especially for shapes with rich details. To speed up the verification process, researchers recently design more powerful local shape descriptors [2], [3], aiming at reducing the number of false positives in the selected correspondences.

Rather than selecting correspondences simply based on the descriptor similarity, mutual interactions among the potential

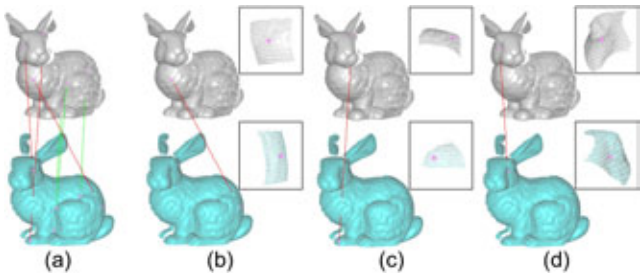

FIGURE 1. (a) Find corresponding features using local shape descriptors, where green and red lines indicate correct and incorrect correspondences. Incorrect correspondences could be generated due to (b) less salient, (c) incomplete (e.g., close to scan boundary), and (d) symmetric shape features.

correspondences can be taken into account for selecting more reliable correspondences. Some researchers propose to group correspondences nearby a seed correspondence based on geometric compatibility [4], [5]. Other researchers reason all potential correspondences simultaneously to find the largest isometry-maintained clusters [6], [7]. All these approaches assume that corresponding features should appear in clusters, and thus ignore isolated ones. Although this assumption could be valid for ideal datasets, it is hard to be satisfied by real scans that are captured under clutter, occlusion, and even noise.

This paper aims at robust and efficient 3D object recognition in cluttered scenes, and develops a framework with three key components to achieve this:

- *A Robust Descriptor:* We propose a novel local shape descriptor, *Signature of Geometric Centroids* (SGC), which is highly descriptive, robust to occlusion and noise, and supports matching feature points near scan boundary [8]. By applying SGC to match scene scan and candidate object models, high-quality potential correspondences (with less false positives) can be obtained.
- *A Correspondence Compatibility Measure.* To support mutual interactions among potential correspondences to find reliable ones, we need to measure geometric compatibility between a pair of potential correspondences. One commonly used compatibility measure is isometry, requiring preserving the point distance. We develop a more powerful compatibility measure by enriching the isometry with three angle-preserving components.
- *A Correspondence Selection Approach:* Observing that all correct correspondences should be compatible with one another while incorrect correspondences can be compatible with others (correct or incorrect ones) very accidently, we propose an effective correspondence selection approach by choosing potential correspondences that have high descriptor similarity and are compatible with an auxiliary correspondence set, which is composed of "less" potential correspondences yet still contains a certain amount of correct correspondences.

We conduct experiments on publicly available datasets [9]–[13] for the descriptor, selection approach, and recognition framework respectively. Experimental results show that SGC descriptor is more robust against noise, varying point density, and distance to scan boundary than state-of-the-art descriptors, while our selection approach is superior than three state-of-the-art selection approaches, especially under challenging scenarios such as high occlusion. Quantitative experiments also demonstrate the efficiency and robustness of our recognition approach.

II. RELATED WORK

LOCAL SHAPE DESCRIPTORS

Local shape descriptors encode the local shape (i.e., support) around a feature point on a given 3D shape into a vector or histogram. By computing and comparing local shape descriptors, potential feature correspondences can be built for two different shapes. Early local shape descriptors are generated by simply accumulating geometric attributes into a histogram such as Spin Images [14], Surface Signatures [15], and 3D shape context (3DSC) [16].

Recently, researchers construct local shape descriptors with relative to a unique local reference frame (LRF); typical descriptors include Signature of Histograms of Orientations (SHOT) [11], Rotational Projection Statistics (RoPS) [3], Local Voxelizer [2]. By constructing a unique LRF for the descriptors, a rigid transform to algin two 3D shapes can be calculated from a single pair of matched descriptors based on aligning the LRFs. Please refer to [17] for more details on state-of-the-art local shape descriptors.

A. SELECTING FEATURE CORRESPONDENCES

By matching local shape descriptors, a large amount of potential correspondences can be generated. Based on the number of candidates considered at a time, existing correspondence selection approaches can be classified into two classes: individual based and group based approaches. Individual based approaches simply select a few correspondence candidates with the highest descriptor similarity [10], [16], [18], or use nearest neighbor similarity ratio [3] to select good candidates that are more unique.

In sharp contrast, group based approaches achieve better performance for selecting reliable correspondences since they consider descriptor similarity of each correspondence candidate, as well as interactions among the candidates. We classify group based approaches into two classes according to whether they employ interactions among a subset or the whole set of candidates.

Local Approaches group a subset of candidate correspondences that are nearby (usually around a seed correspondence) and eliminate outliers using geometric consistency [18]. Chen and Bhanu [4] partitioned all matched point pairs into different groups by using the isometry constraint, and selected the group with the largest size as reliable correspondences. Aldoma *et al.* [5] refined the approach [4] by adding a subsequent RANSAC on each generated group to remove spurious correspondences in the group. Buch *et al.* [19] selected reliable correspon-

dences using a two-step approach: filter the correspondence candidates by locally applying the geometric consistency constraint; and then globally apply covariant constraints to the filtered candidates, requiring compatible transforms hypothesized by the candidates.

Global Approaches consider all candidate correspondences simultaneously while enforcing various constraints such as geometric compatibility. Leordeanu and Hebert [6] built an adjacency matrix for a graph whose nodes represent candidate correspondences and the weights on the edges represent pairwise agreements between the candidates, and main correspondence cluster is found by computing the principal eigenvector of the matrix. Torresani *et al.* [20] established correspondences between sparse image features based on their appearance and spatial arrangement by solving an energy minimization problem using graph matching techniques. Rodolà *et al.* [7] developed a game-theoretic framework for selecting correspondences that satisfy global consistency constraints.

Unlike above approaches that enforce geometric compatibility constraint within candidate correspondences, we identify correct correspondences from the candidates by measuring their compatibility with another auxiliary correspondence set. We do not assume that correct correspondences should appear in cluster, and thus are able to select isolated correspondences also. In addition, we speed up our selection approach by taking advantage of recent development on approximate nearest neighbor algorithms.

B. 3D OBJECT RECOGNITION

3D object recognition [1] in the literature can be broadly classified into two categories: global- and local-based approaches. Global-based approaches such as shape distribution [21] and viewpoint feature histogram [22] try to encode geometric properties of the whole shape. However, they require a priori segmentation on the object from the scene and ignore local shape details, making them not suitable for recognizing partially occluded objects.

In contrast, local-based approaches [23], [24] employ local geometric shape features to find correspondences, thus are more suitable for recognition in cluttered scenes. Johnson and Hebert [18] generated feature correspondences by matching compressed spin images and grouping geometrically consistent correspondences for model verification. Frome *et al.* [16] represented sampled surface patches using 3D shape context (3DSC) descriptors and performed the recognition by aggregating the descriptor distances to make a choice as to which model is the best match to the scene. Mian *et al.* [10] developed a fully automatic recognition approach by matching tensor descriptors to generate a model hypothesis, which is further verified by checking if the model aligns accurately with an object in the scene. Guo *et al.* [3] constructed a rotational projection statistics (RoPS) descriptor with a unique LRF and applied the descriptor for recognition, where a single pair of matched descriptors votes for a model hypothesis by using the LRF. Guo *et al.* [25] later developed another 3D object recognition algorithm based on hierarchical matching of Tri-Spin-Image features. We refer readers to an excellent survey [1] for more details on state-of-the-art 3D object recognition algorithms using local shape descriptors.

Compared with above local-based approaches, our 3D object recognition approach improves the effectiveness when generating and selecting candidate correspondences. First, we propose a descriptive and robust SGC descriptor for generating high-quality candidate correspondences. Second, we develop an effective correspondence selection approach by enforcing geometric compatibility between each candidate correspondence and another auxiliary correspondence set.

III. OVERVIEW OF OUR RECOGNITION PIPELINE

3D object recognition aims at identifying objects that are present in a scene and recover their poses. To achieve this, local shape descriptors are employed to describe local shape features on the scene and object models, and potential feature correspondences can be built by matching the descriptors and computing their similarities. Next, a few reliable correspondences can be selected from the potential correspondences to vote for models that are likely to appear in the scene. Lastly, a model verification step is adopted to check if the model does really appear in the scene by checking the alignment between the transformed model and the scene.

A. DESCRIBE LOCAL SHAPE

Originally captured scans are mostly represented as point clouds, and contain nuisances such as noise and occlusion. To describe shape of original scans well, we propose an SGC descriptor that takes point cloud data as input and is robust against the various nuisances. We devise the SGC descriptor by voxelizing the local shape within a uniquely defined LRF and concatenating geometric centroid and point density features extracted from each voxel. When comparing two SGC descriptors, we only consider corresponding voxels that are both non-empty, thus supporting matching incomplete local shape such as those close to scan boundary (Section IV).

B. CONSTRUCT A CANDIDATE CORRESPONDENCE SET

Given a model M and a scene S (Figure 2(a)), feature points are uniformly sampled on the whole surface of M and S respectively. Note that key-point detection techniques [26] can also be adopted here to detect more salient local shapes. Next, we construct an SGC descriptor for the support around each feature point (Figure 2(b)). We compare each descriptor of S with those in M and compute similarity scores. A feature point on S and its closest feature point (with the highest similarity score) on M are considered as a candidate correspondence. By summarizing the candidate correspondence for every feature point on S, we generate a candidate correspondence set $\mathcal{G}$ (Figure 2(c)).

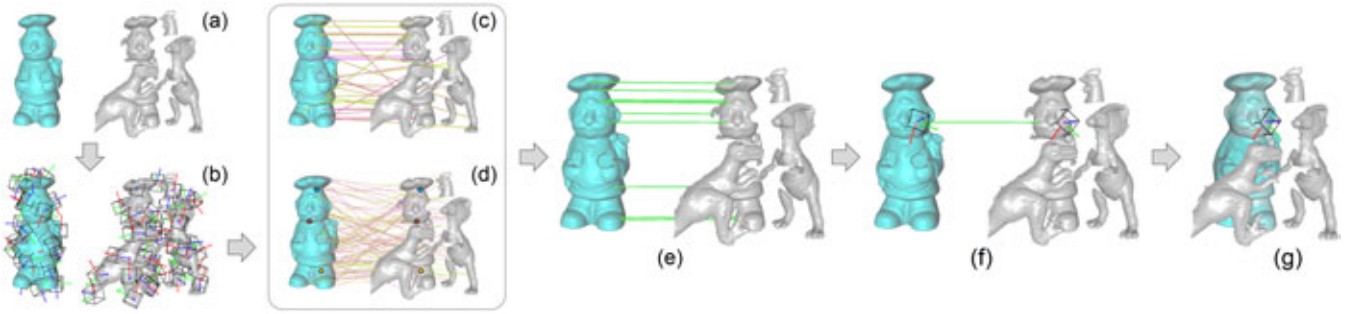

FIGURE 2. Our 3D object recogntion pipeline. (a) A model (Chef) and a scene; (b) constructed SGC descriptors for sampled feature points; (c) set $\mathcal{G}$ with candidate correspondences (in thicker line), where high and low descriptor similarities are colored in yellow and purple respectively; (d) set $\mathcal{A}$ with auxiliary correspondences; (e) select K correspondences from set $\mathcal{G}$, where green lines represent correct correspondences; (f) a single pair of corresponding features (i.e., one correspondence) votes for a model hypothesis, where model-to-scene transform is estimated based on aligning the LRFs; (g) verify the hypothesized model.

C. CONSTRUCT AN AUXILIARY CORRESPONDENCE SET

For a feature point on S, besides the closest feature point on M, the following L closest feature points on M also have a reasonable chance to be the corresponding point. However, these potential correspondences are usually considered as redundant and discarded directly for efficiency. Rather, this paper makes use of these ''less'' potential correspondences to generate an auxiliary correspondence set $\mathcal{A}$, which usually contains a certain amount of correct correspondences (three pairs of corresponding points are manually labelled in Figure 2(d)).

D. SELECT RELIABLE CORRESPONDENCES

We adopt the auxiliary set $\mathcal{A}$ to vote for the candidate correspondences in $\mathcal{G}$ based on the observation that correct correspondences in $\mathcal{G}$ should be geometrically compatible with the correct ones in $\mathcal{A}$. In particular, we develop a correspondence compatibility measure by enriching the well known isometry with three angle-preserving components (Section V). By this, we can evaluate each correspondence in $\mathcal{G}$ by using its original descriptor similarity and its compatibility measure with the correspondences in $\mathcal{A}$ (or a portion of $\mathcal{A}$ for speeding up). We further select the top K candidate correspondences received the best evaluation for model verification (Section VI), which are likely to be correct correspondences (Figure 2(e)).

E. VERIFY THE HYPOTHESIZED MODEL

Taking advantage of the unique LRF in the SGC descriptor, each candidate correspondence can vote for a model hypothesis and generate a model-to-scene transform by aligning associated LRFs. We then evaluate the model hypothesis according to the alignment between the transformed model and the scene (Section VII). A candidate correspondence (and its hypothesized model) is accepted if the overlap is larger than a predefined threshold.

IV. SIGNATURE OF GEOMETRIC CENTROIDS DESCRIPTOR

This section presents the construction an SGC descriptor, and the scheme to compare a pair of SGC descriptors. More technical details can be found in [8].

A. SGC CONSTRUCTION

Given a support around a feature point, an SGC descriptor assigns the feature point a vector that encodes the spatial and geometric characteristics of the support with relative to a uniquely defined LRF.

1) LRF CONSTRUCTION

Given a feature point p on a scan and a radius r, a local support is defined by intersecting the scan with a sphere centered at p with radius r. Taking this support as input, we construct a unique LRF based on principal component analysis on the support following [11], see Figure 3(a).

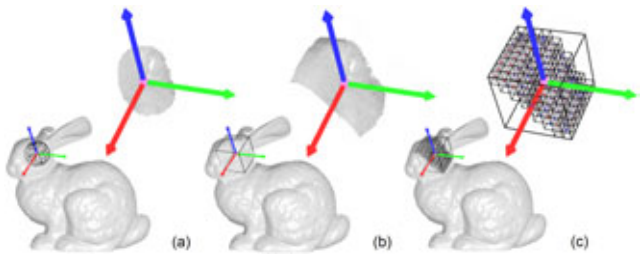

FIGURE 3. Constructing an SGC descriptor. (a) Construct a unique LRF from a spherical support centered at a feature point (in pink); (b) segment a cubical support centered at the point and aligned with the LRF; (c) voxelize the support and extract centroid features from non-empty voxels; the centroid color indicates point density in the voxel, where small and large densities are colored in blue and red respectively.

Given a support S_p around a feature point p, we construct a bounding cubical volume of S_p that is aligned with the LRF and partition the cubical volume into regular bins (i.e., voxels) similar as [2], see Figure 3(b&c). Note that edges of the cubical volume have a length of $2R$, where $R \geq r$. we choose $R = r = 20\ pr$ as a tradeoff between the descriptiveness and sensitivity to occlusion, where pr denotes the point cloud resolution (i.e., average shortest distance among neighboring points in the scan).

2) CONSTRUCT THE DESCRIPTOR

We divide the cubical volume evenly into $k \times k \times k$ bins (i.e., voxels) with the same sizes, see Figure 3(c). We choose

$k = 8$ as a tradeoff between the descriptiveness and efficiency since a larger k increases the descriptiveness and computational cost simultaneously. For each voxel $\mathrm{V_i}$, we identify all O_i points staying within the voxel and then calculate the centroid (X_i, Y_i, Z_i) for the points. Note that, the position of the centroid is relative to the minimum corner of $\mathrm{V_i}$ in the LRF. For each non-empty voxel $\mathrm{V_i}$, we save the extracted feature using four values as (X_i, Y_i, Z_i, O_i). Otherwise, we save the feature as (0,0,0,0). An SGC descriptor for point p is generated by concatenating all these values assigned for each voxel.

B. COMPARING SGC DESCRIPTORS

Ideally, SGC descriptors generated for two corresponding points in different scans should be exactly the same. However, due to variance of sampling, noise and occlusion, the two descriptors usually have a certain amount of difference. We develop a new scheme for comparing two SGC descriptors.

When constructing an SGC descriptor, most of the voxels are likely to be empty (see again Figure 3(c)). We classify each pair of corresponding voxels into three cases: 1) empty voxel vs empty voxel; 2) non-empty voxel vs empty voxel; and 3) non-empty voxel vs non-empty voxel. In all three cases, only case 3 should contribute to computing a similarity score between two descriptors. Thus, to compare two SGC descriptors quantitatively, we accumulate a similarity score for every pair of corresponding voxels that are both non-empty.

In detail, we denote two SGC descriptors as $\mathrm{D_m}$ and $\mathrm{D_n}$. The similarity between the i-th voxel of $\mathrm{D_m}$, $\mathrm{V_m^i}$, and the i-th voxel of $\mathrm{D_n}$, $\mathrm{V_n^i}$, is defined as:

$$s(\mathrm{V_m^i}, \mathrm{V_n^i}) = \begin{cases} \ln \dfrac{O_m^i O_n^i}{\|C_m^i - C_n^i\|^2 + \epsilon}, & \text{for } O_m^i > 0 \text{ and } O_n^i > 0 \\ 0 & \text{for } O_m^i = 0 \text{ or } O_n^i = 0 \end{cases} \tag{1}$$

where O_m^i and O_n^i represent the number of points in $\mathrm{V_m^i}$ and $\mathrm{V_n^i}$ respectively, while C_m^i and C_n^i represent the centroid of $\mathrm{V_m^i}$ and $\mathrm{V_n^i}$ respectively. Here we directly employ the number of points in each voxel to represent its point density as all voxels have the same size. The formula can be explained as follows. Whenever $\mathrm{V_m^i}$ and/or $\mathrm{V_n^i}$ are empty (i.e., $O_m^i = 0$ or $O_n^i = 0$), $s(\mathrm{V_m^i}, \mathrm{V_n^i}) = 0$. Otherwise, when two corresponding voxels contain similar local shape, their centroids should be close to each other, making $s(\mathrm{V_m^i}, \mathrm{V_n^i})$ large. When O_m^i and/or O_n^i are large, $s(\mathrm{V_m^i}, \mathrm{V_n^i})$ is large also as the estimated centroid(s) are more accurate.

The overall similarity score between $\mathrm{D_m}$ and $\mathrm{D_n}$ can be obtained by accumulating the similarity value for every pair of corresponding voxels:

$$S(\mathrm{D_m}, \mathrm{D_n}) = \sum_{i=1}^{k \times k \times k} s(\mathrm{V_m^i}, \mathrm{V_n^i}) \tag{2}$$

This similarity measure does not require complete supports for the comparison, thus can match local shape that is partially occluded or close to scan boundary.

V. CORRESPONDENCE COMPATIBILITY MEASURE

This section presents a geometric compatibility measure for two (“less”) potential correspondences, which integrates four compatibility components.

A. COMPATIBILITY COMPONENTS

Given a data scan S_d and a reference scan S_r (assume a rigid transform), we denote a potential feature correspondence between a point P_i on S_d and a point P_j on S_r as follows:

$$\mathcal{H}_{i,j} = (P_i, P_j, N_i, N_j, S_{i,j}) \tag{3}$$

where N_i and N_j are the associated normal of P_i and P_j respectively, and $S_{i,j}$ is the descriptor similarity score. Our compatibility measure consists of four components: one distance-preserving component (i.e., isometry) and three angle-preserving components.

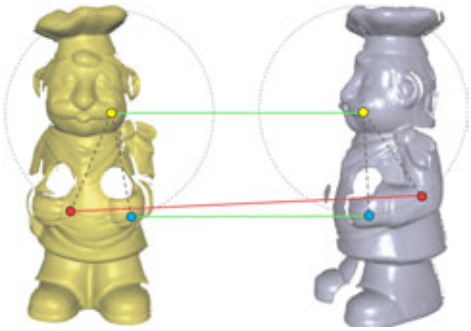

FIGURE 4. Ambiguity of isometry constraint. Two pairs of matched features that fulfill the isometry constraint may (yellow and blue point pairs) or may not (yellow and red point pairs) be both corresponding features.

1) ISOMETRY COMPONENT

Isometry component requires that if two correspondences $\mathcal{H}_{i,j}$ and $\mathcal{H}_{k,l}$ are both correct, then the Euclidean distance $||P_i - P_k||$ on the data scan S_d should be close to the distance $||P_j - P_l||$ on the reference scan S_r, see Figure 4. Isometry can be measured by calculating the difference between the two distances [5],

$$d_1(\mathcal{H}_{i,j}, \mathcal{H}_{k,l}) = \big|\, ||P_i - P_k|| - ||P_j - P_l|| \,\big| \tag{4}$$

However, the isometry constraint has inherent ambiguity, which means that two pairs of features fulfill the isometry constraint need not be both corresponding features, see again Figure 4. Thus, we propose another three angle-preserving measures to enrich it.

2) ANGLE-PRESERVING COMPONENTS

For two correspondences $\mathcal{H}_{i,j}$ and $\mathcal{H}_{k,l}$, angle-preserving constraints require that three angles computed from (P_i, N_i, P_k, N_k) on the data scan S_d should be close to those

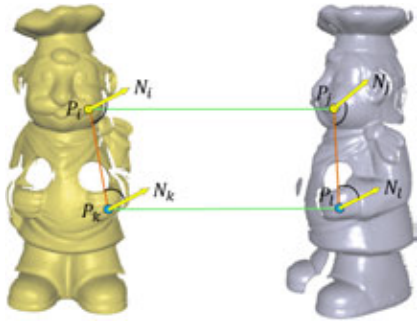

FIGURE 5. Angle-preserving components. For two pairs of matched points (yellow and blue point pairs) on two different scans, three angles on the data scan (left) should be close to those on the reference scan (right), respectively.

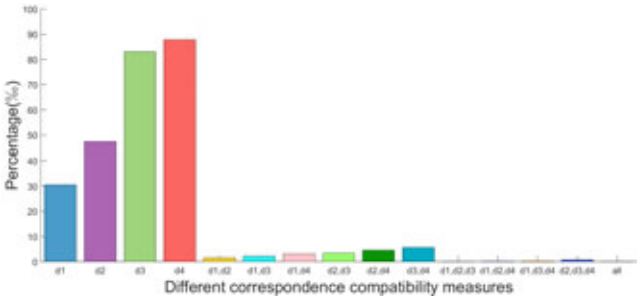

FIGURE 6. The percent of incorrect correspondences that are considered compatible with correct correspondences by different compatibility measures, from left to right, isometry, each angle-preserving measure, and their combinations (our measure is on the right most).

computed from (P_j, N_j, P_l, N_l) on the reference scan S_r, see Figure 5. Specifically, the three angles (e.g., on S_d) are:

- Angle between N_i and N_k;
- Angle between N_i and $\overrightarrow{P_iP_k}$;
- Angle between N_k and $\overrightarrow{P_kP_i}$.

By calculating these three angles on S_d and S_r respectively and comparing each pair of them, we can evaluate how the angle-preserving constraints are satisfied by using the following three measures:

$$d_2(\mathcal{H}_{i,j}, \mathcal{H}_{k,l}) = \left|\angle(N_i, N_k) - \angle(N_j, N_l)\right| \quad (5)$$

$$d_3(\mathcal{H}_{i,j}, \mathcal{H}_{k,l}) = \left|\angle(N_i, \overrightarrow{P_iP_k}) - \angle(N_j, \overrightarrow{P_jP_l})\right| \quad (6)$$

$$d_4(\mathcal{H}_{i,j}, \mathcal{H}_{k,l}) = \left|\angle(N_k, \overrightarrow{P_kP_i}) - \angle(N_l, \overrightarrow{P_lP_j})\right| \quad (7)$$

Note that our compatibility constraints are inspired by the point pair feature developed in [24]. However, due to the limited descriptiveness, direct shape matching with point pair features is not comparable with using SGC descriptors in terms of effectiveness. Rather, this paper employs SGC to find a set of good correspondence candidates and further identify true positives in the candidates by enforcing these compatibility constraints.

B. COMPATIBILITY MEASURE

Two correspondences, $\mathcal{H}_{i,j}$ and $\mathcal{H}_{k,l}$, are more likely to be geometrically compatible if they fulfill the isometry and angle preserving constraints simultaneously. Thus we define the compatibility measure as follows:

$$C(\mathcal{H}_{i,j}, \mathcal{H}_{k,l}) = \exp\Big(-\frac{d_1(\mathcal{H}_{i,j}, \mathcal{H}_{k,l})^2}{2{\sigma_1}^2} - \frac{d_2(\mathcal{H}_{i,j}, \mathcal{H}_{k,l})^2}{2{\sigma_2}^2} - \frac{d_3(\mathcal{H}_{i,j}, \mathcal{H}_{k,l})^2}{2{\sigma_3}^2} - \frac{d_4(\mathcal{H}_{i,j}, \mathcal{H}_{k,l})^2}{2{\sigma_4}^2}\Big) \quad (8)$$

where σ_1, σ_2, σ_3 and σ_4 are the weights to control the impact of isometry and three angle-preserving components respectively. In our experiments, we set $\sigma_1 = 10pr$ and $\sigma_2 = \sigma_3 = \sigma_4 = \frac{1}{6}\pi$. Note that $C(\mathcal{H}_{i,j}, \mathcal{H}_{k,l})$ is large only when all the four constraints are satisfied, and is close to zero when some constraints are violated.

To evaluate the effectiveness of our compatibility measure, we compare it with isometry, each angle-preserving measure, and their combinations. We generate a set of correct correspondences and another set of incorrect correspondences, and count how many incorrect correspondences are considered as compatible with correct correspondences by using each of these measures. Results in Figure 6 show that although isometry is more effective than each of the three angle-preserving components, our measure that combines all the four components performs the best (i.e., zero false positive).

VI. CORRESPONDENCE SELECTION APPROACH

This section presents our correspondence selection approach that leverages the aforementioned candidate correspondence set $\mathcal{G}$ (Section III), auxiliary correspondence set $\mathcal{A}$ (Section III), and correspondence compatibility measure (Section V).

A. CORRESPONDENCE SELECTION VIA AUXILIARY SET VOTING

Selecting corresponding features based on geometric compatibility has been investigated in the literature, where the general idea is to find the largest isometry preserved clusters [6], [7] as the corresponding features. However, these existing approaches have two limitations. First, under challenging scenarios (e.g., high occlusion), corresponding features may appear isolated rather than in cluster, making these approaches fail. Second, a large amount of false positives can be selected by simply enforcing the isometry constraint due to its ambiguity (see again Figure 4).

Our selection approach resolves the above two limitations by developing an auxiliary set voting scheme that enforces the more powerful compatibility measure between $\mathcal{G}$ and $\mathcal{A}$. Since $\mathcal{A}$ also contains a certain amount of corresponding features, each correct correspondence in $\mathcal{G}$ should be geometrically compatible with all the correct ones in $\mathcal{A}$. Thus, we can vote each $\mathcal{H}_{i,j}$ in $\mathcal{G}$ using all the correspondences in $\mathcal{A}$, and evaluate $\mathcal{H}_{i,j}$ using Equation 9 that considers both descriptor similarity and geometric compatibility:

$$S'_{i,j} = S_{i,j} + \omega \sum_{\mathcal{H}_{k,l} \in \mathcal{A}} S_{k,l} \times C(\mathcal{H}_{i,j}, \mathcal{H}_{k,l}) \quad (9)$$

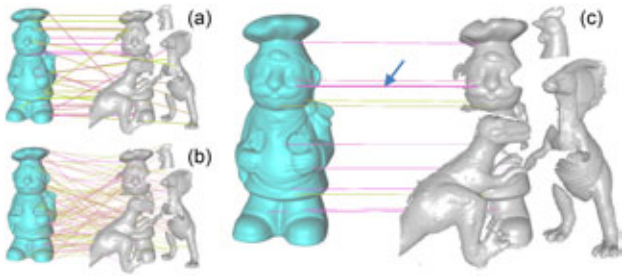

FIGURE 7. (a) Candidate correspondences; (b) auxiliary correspondences; (c) a correct candidate correspondence with a low similarity score is voted by compatible auxiliary correspondences, thus improving its evaluation. High and low descriptor similarities are colored in yellow and purple respectively.

By this, we can perform correspondence selection based on this new evaluation measure, see Figure 7. Note that we weight each compatibility score $C(\mathcal{H}_{i,j}, \mathcal{H}_{k,l})$ using the similarity score of $\mathcal{H}_{k,l}$ (i.e., $S_{k,l}$) such that correspondences in $\mathcal{A}$ with a higher similarity score can contribute more for the voting.

B. SPEED UP THE SELECTION APPROACH

To ensure a good selection performance, the number of correspondences in $\mathcal{A}$ should be as large as possible such that more correct correspondences are contained. However, the large cardinality of $\mathcal{A}$ makes the exhaustive comparison between $\mathcal{G}$ and $\mathcal{A}$ extremely slow, violating our goal of improving recognition efficiency. To speed up our selection process, we propose to compare each correspondence $\mathcal{H}_{i,j}$ in $\mathcal{G}$ with a small subset of $\mathcal{A}$ whose elements are (approximately) most compatible with $\mathcal{H}_{i,j}$.

For each correspondence $\mathcal{H}_{i,j}$ in $\mathcal{G}$, to fast retrieve a small set of elements in $\mathcal{A}$ that are most compatible with $\mathcal{H}_{i,j}$, the common approach is to index all elements in $\mathcal{A}$ based on an element distance measure. For example, for a 3D point cloud, we can build a k-dimensional tree (k-d tree) to search k-nearest neighbors for a query point based on the Euclidean distance measure. However, our developed correspondence compatibility measure (i.e., Equation 8) is non-Euclidean and does not satisfy the triangle inequality property, making these classical techniques infeasible.

To make our problem tractable, we relax the problem of finding exact nearest neighbors by searching approximate nearest neighbors instead, and take advantage of recent development on approximate nearest neighbor algorithms such as small world graphs [27]. In particular, we find Hierarchical Navigable Small World graphs (HNSW) [28], [29] are able to handle arbitrary metric spaces (e.g., not limited to Euclidean metric space) and are more universal. Thus, we build an HNSW graph for all correspondences in $\mathcal{A}$, which is fast since element insertion requires only local information of the graph structure. To simplify the distance computation (i.e., inverse of compatibility measure), we formulate the distance measure between two correspondences $\mathcal{H}_{i,j}$ and $\mathcal{H}_{k,l}$ as:

$$D(\mathcal{H}_{i,j}, \mathcal{H}_{k,l}) = -\log C(\mathcal{H}_{i,j}, \mathcal{H}_{k,l}) \tag{10}$$

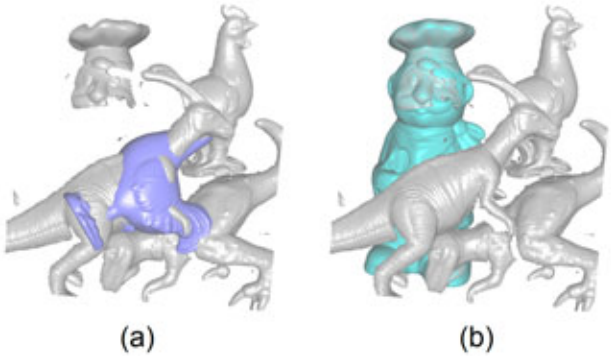

FIGURE 8. (a) A falsely hypothesized model gets a model-to-scene overlap of 9.12%; (b) a highly occluded model gets a model-to-scene overlap of 8.73%.

For $\mathcal{H}_{i,j}$ in $\mathcal{G}$, we can efficiently retrieve a small set of auxiliary correspondences $\mathcal{A}(i,j)'$ in $\mathcal{A}$ that are (approximately) most compatible with $\mathcal{H}_{i,j}$ by searching it across the HNSW graph. Then instead of comparing $\mathcal{H}_{i,j}$ with all auxiliary correspondences in $\mathcal{A}$, we compute the compatibility measure only with those in $\mathcal{A}(i,j)'$. Thus, the final score $S'_{i,j}$ can be redefined as:

$$S'_{i,j} = S_{i,j} + \omega \sum_{\mathcal{H}_{k,l} \in \mathcal{A}(i,j)'} S_{k,l} \times C(\mathcal{H}_{i,j}, \mathcal{H}_{k,l}) \tag{11}$$

By re-ranking the correspondences in $\mathcal{G}$ based on $S'_{i,j}$, we obtain a reordered list of candidate correspondences denoted as $\mathcal{G}'$. We further select the top K correspondences in $\mathcal{G}'$ as the input of model verification stage.

VII. MODEL VERIFICATION

Taking the K selected correspondences as input, we try each of them iteratively following the ranking order to generate a hypothesized model and perform the model verification. Our model verification stage is fast due to two reasons: 1) the selected K correspondences are likely to be correct; 2) one correct correspondence is sufficient to complete the model verification process.

Each correspondence between a scene scan and a candidate model hypothesizes that the model appears in the scene. Taking advantage of the unique LRF of our SGC descriptor, each correspondence can generate a model-to-scene transform by aligning associated LRFs. We further refine the transform by applying the ICP algorithm [30] for a few iterations. Ideally, correctly hypothesized model should be seamlessly aligned with the scene scan. However, due to noise and/or varying point density in the scans, the model-to-scene alignment cannot be perfect. Thus, we estimate the overlap between the transformed model M' and the scene S as follows: first, find all point-to-point correspondences by checking if the distance between a point on M' and a point on S is sufficiently small; and then compute the overlap ratio as the number of corresponding points divided by the total number of points in M'.

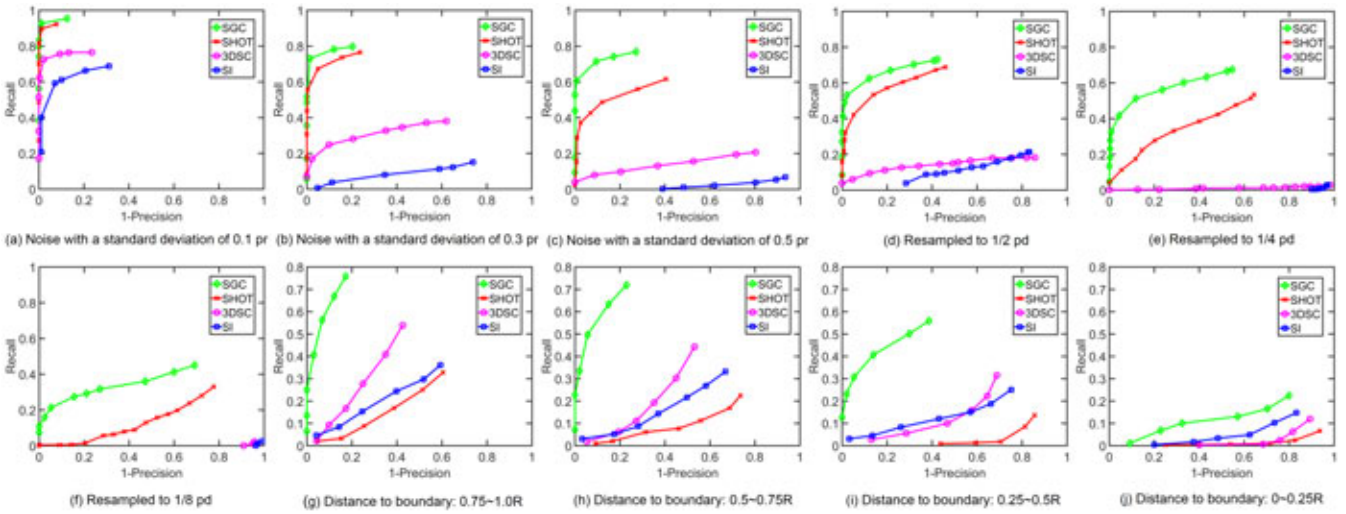

FIGURE 9. RP curves of SGC and three state-of-the-art descriptors in the presence of (a-c) noise, (d-f) point cloud downsampling, and (g-j) distance to scan boundary.

This overlap ratio (i.e., visible proportion of model in the scene) has been commonly adopted by existing methods [2], [3], [10] for model verification. However, the threshold of overlap ratio is hard to set since a large threshold could accept incorrect models that are aligned with the scene accidently while a small threshold will reject highly occluded models (see Figure 8).

To overcome this, we propose a new thresholding strategy. We compute the variance of the nearest point distances between the transformed model and the scene at the overlap region. And we observe that these distances should have a small variation when the hypothesized model is correct and thus well aligned with the scene. Specifically, we choose a low threshold for the overlap ratio together with a strict variation threshold for verifying highly occluded objects successfully. By this, we can reject the case in Figure 8(a) while accept the case in Figure 8(b) using fixed thresholds.

When the hypothesized model is accepted, we segment out the overlap region from the scene scan and discard all the remaining hypotheses that are generated by the feature points located at the overlap region. This process continues until most surface in the scene scan are successfully segmented or no hypothesis is left.

VIII. EXPERIMENTS RESULTS

We implement our methods in C++ and execute it on a desktop PC with an Intel Xeon E3-1230 v3 CPU (3.4GHz, 4 cores) and 8GB memory without using any parallel computing techniques. We build and search HNSW graph for correspondence selection by employing the Non-Metric Space Library [31]. To speed up descriptor comparison between models and a scene, we build another HNSW graph off-line to index sampled descriptors from all candidate models.

A. EVALUATE SGC DESCRIPTOR

We compare our SGC with three state-of-the-art descriptors: Spin image (SI) [14], 3DSC [16] and SHOT [11] on two publicly available datasets: the Bologna dataset [11] to evaluate the robustness against noise and varying point density (pd) and the UWA recognition dataset [10] to evaluate the robustness against the distance to scan boundary. For a fair comparison, we set the support radius $R = 20pr$ for all descriptors while all other parameters are followed the settings in their original works.

The comparison is evaluated using the criterion of recall versus 1-precision curve (RP curves) [32], see Figure 9. The plots shows that SGC is very robust to noise, varying point density and the distance to scan boundary, and outperforms the other three descriptors on a large margin. SHOT is also robust to noise and varying point density, but is very sensitive when the feature points are close to scan boundary. On the contrary, 3DSC and SI perform well when the feature points locate slightly near boundary. We refer readers to [8] for more details on this experiment.

B. EVALUATE CORRESPONDENCE SELECTION APPROACH

We evaluate our correspondence selection approach on three publicly available datasets: SHOT occlusion dataset [12], Bologna dataset [11], and UWA modeling dataset [9]. Note that, in each scene of the SHOT occlusion dataset [12], one of the 4 models appears at different levels of occlusion and clutter.

1) TUNING PARAMETER

As our aim is to leverage the correct correspondences in the auxiliary set $\mathcal{A}$ to encourage true positives in $\mathcal{G}$, the number of auxiliary correspondences is critical for our selection performance. A larger auxiliary set will contain more correct correspondences, making the true positives in $\mathcal{G}$ have a higher chance to be voted, yet it also requires more computation cost. Thus, this experiment focuses on tuning the parameter L for constructing the auxiliary set $\mathcal{A}$. We employ recall@500 as the metric for quantitative comparison which

tells the ratio of true positives that can be retrieved within the selected 500 candidates over the total number of correct correspondences.

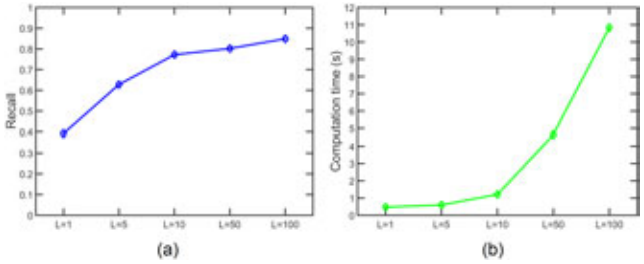

FIGURE 10. (a) Recall of the first 500 correspondences with respect to L; (b) computation time with respect to L.

We conduct this experiment on the UWA modeling dataset [9]. In detail, we randomly generate 2000 feature points on the data scan S_d and reference scan S_r respectively, obtaining $M = 2000$ candidate correspondences and $N = LM$ auxiliary ones. Figure 10(a) presents the recall of the first 500 correspondences with relative to the increasing L, showing that the selection performance improves significantly when L increases from 1 to 10. Figure 10(b) presents the corresponding computation time, showing that it takes around $1sec$ when $L \leq 10$. Therefore, we select $L = 10$ as a tradeoff between effectiveness and computation cost.

2) COMPARISON WITH STATE-OF-THE-ARTS

This section evaluates the efficiency and robustness of our selection approach by comparing it with three state-of-the-art correspondence selection approaches. To ensure a fair comparison, we generate exactly the same set of potential correspondences using the SGC descriptor as the input of the four selection approaches.

We first describe the three state-of-the-art selection approaches briefly as follows:

- *Similarity score (SS).* The most common selection approach is to rank all correspondences simply according to the descriptor similarity score.
- *Nearest neighbor similarity ratio (NNSR).* One variation of similarity score approach is to compute a nearest neighbor similarity ratio [33] to rank all correspondences, i.e., the ratio between the second maximum similarity and the maximum similarity.
- *Spectral technique (ST).* Spectral technique [6] is a joint method that builds an adjacency matrix, considering descriptor similarity score and isometry compatibility measure simultaneously. By conducting eigen decomposition on the matrix, the confidence of each correspondence can be obtained as the corresponding value in the principal eigenvector.

a: ROBUSTNESS TO NOISE

To evaluate the performance of the four selection approaches under noise, we generate scenes by adding Gaussian noise with a standard deviation of 50% pr on the complete models in the Bologna dataset [11] (Figure 11(b)). Figure 11(c) visualizes the correctness of 500 selected correspondences by the four approaches, showing that top-ranked correspondences by our Auxiliary Set Voting (ASV) approach are mostly correct, followed by those of ST. This indicates that the correspondence compatibility constraint enforced by ASV is able to reject false positives generated when matching SGC descriptors, even under noise.

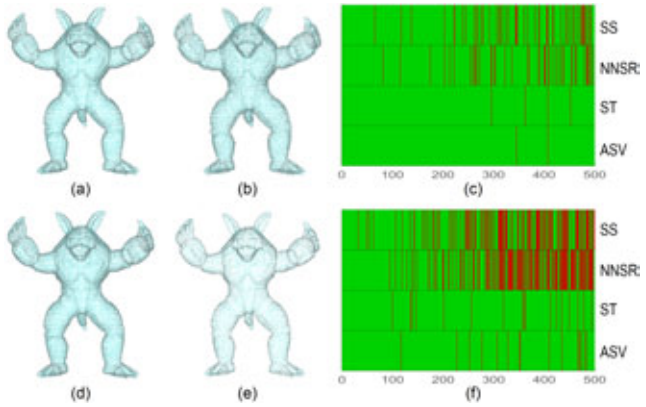

FIGURE 11. Top: Visualizing the correctness of 500 selected correspondences (c) generated between a model (a) and a noisy scene (b). Bottom: Visualizing the correctness of 500 selected correspondences (d) generated between a model (e) and a downsampled scene (f). Correct and incorrect correspondences are visualized in green and red respectively. Top-ranked correspondences are on the left side of the visualizations (c&f), which usually have a higher chance to be correct.

b: ROBUSTNESS TO VARYING POINT DENSITY

To evaluate the performance of the four approaches under varying point density, we generate scenes by resampling the models down to 50% of their original point density (Figure 11(e)). Visualization results for 500 selected correspondences are shown in Figure 11(f). It shows that SS and NNSR select a large number of incorrect correspondences, indicating that the similarity score of SGC descriptors becomes less confident when the scene scan becomes sparse. Thanks to our selection scheme based on auxiliary set voting, ASV outperforms the other three approaches, and is slightly better than ST.

c: ROBUSTNESS TO OCCLUSION

To evaluate the performance of the four approaches under occlusion, we select the SHOT occlusion dataset [12] and group scenes into four categories according to their levels of occlusion, i.e., (50, 60 %], (60 %, 70 %], (70 %, 80 %], and (80 %, 90 %].

Some example selection results are presented in Figure 12(a-d). Selected correspondences by ASV are mostly correct under all levels of occlusion. Although ST performs comparably as ASV under low levels of occlusion (top two rows in Figure 12), it fails completely for highly occluded scenes. This is because ST tries to find large

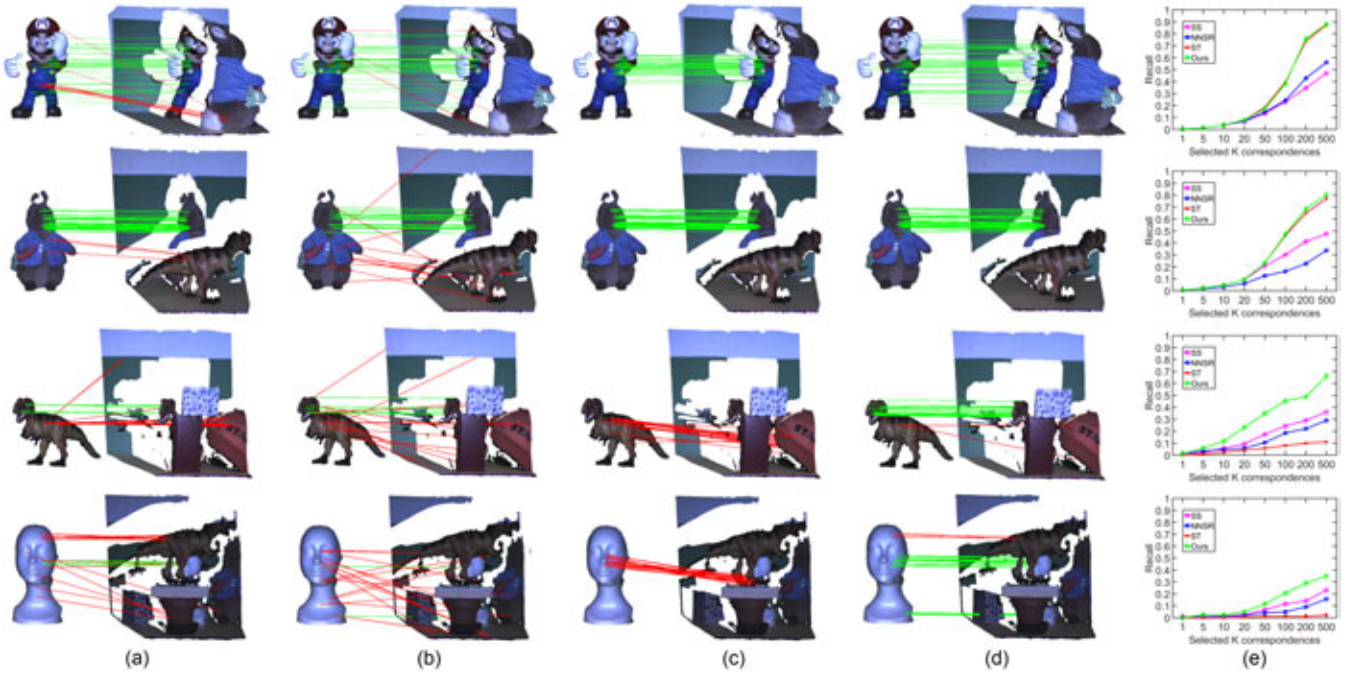

FIGURE 12. Selected correspondences under different levels of occlusion by: (a) SS, (b) NNSR, (c) ST, and (d) ASV. (e) Quantitative comparison of selection performance using CR curves. The occlusion levels from top to bottom are: (50, 60 %], (60 %, 70 %], (70 %, 80 %], and (80 %, 90 %].

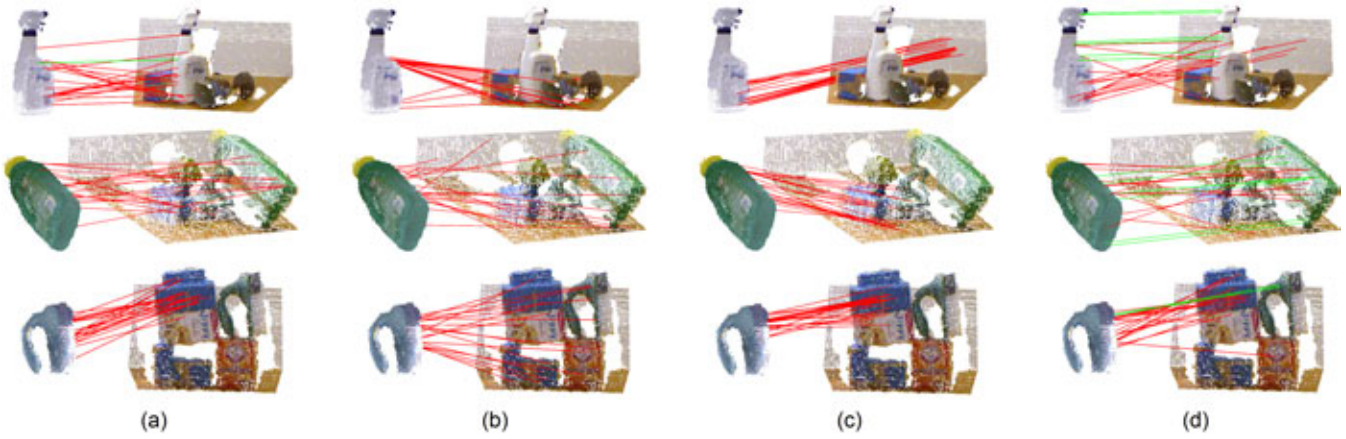

FIGURE 13. Selected correspondences under challenging selection scenarios by: (a) SS, (b) NNSR, (c) ST, and (d) ASV.

isometry-maintained clusters, which do not exist between highly occluded scenes and models. One interesting observation is that NNSR is slightly better than SS under low levels of occlusion while the comparison result is reversed when occlusion becomes severe. This can be explained by the fact that NNSR prefers to select ''distinct" correspondences. Under low levels of occlusion, ''distinct" correspondences are likely to appear at distinct object regions with rich shape features. But, under high levels of occlusion, these distinct object regions are mostly occluded and do not appear in the scene.

We further employ Cumulative Recall Curves (CR curves) as the metric for quantitative comparison, which measures the recall of selected K correspondences. Figure 12(e) presents

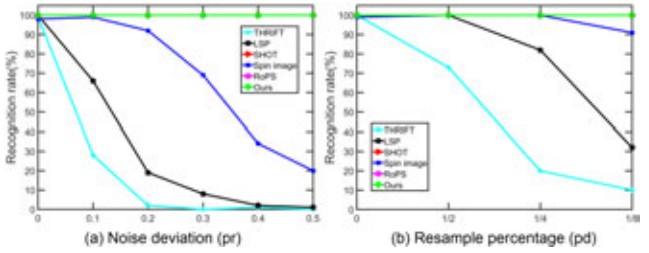

FIGURE 14. Recognition rates with respect to different levels of (a) noise; and (b) varying point density. Note that the curves of SHOT and RoPS coincide with ours and are occluded.

the CR curves, showing that ASV performs best under all levels of occlusion, and its recall of 500 selected correspondences achieves nearly 40% even under high occlusion

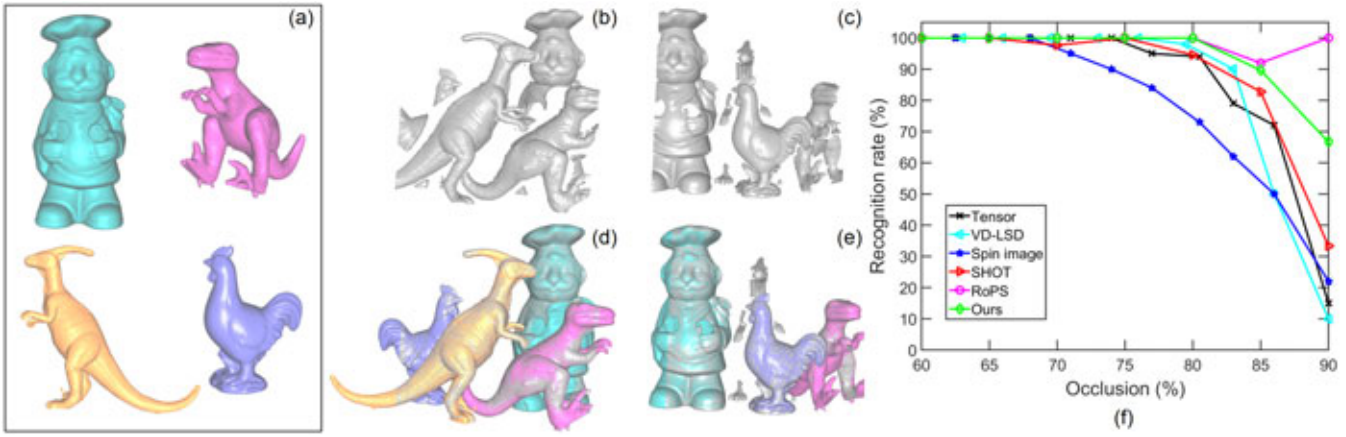

FIGURE 15. Recognition results on the UWA Dataset. (a) Four candidate models: Chef, T-rex, Parasaurolophus, and Chicken (left to right, and then top to bottom). (b,c) Two examples scenes. (d,e) The corresponding recognition results; all objects in the scenes are correctly recognized (see the superimposed models), except the Parasaurolophus in (e) due to the high occlusion. (f) The recognition rate with respect to occlusion.

of (80%, 90%]. ST achieves a comparable performance as ASV under low levels of occlusion(top two rows in Figure 12). Yet, ST performs worse when occlusion becomes severe, and fails completely when the occlusion level exceeds 80%.

d: CHALLENGING SELECTION SCENARIOS

We also conduct correspondence selection experiments on the Clutter dataset [13], which combines noise, occlusion and featureless objects (Figure 13). Under such challenging scenarios, all other three approaches mostly fail (i.e., few green lines in Figure 13(a-c)). Thanks to our auxiliary set voting scheme, our approach is still able to select a small set of correct correspondences (green lines in Figure 13(d)).

C. EVALUATE 3D OBJECT RECOGNITION APPROACH

This section evaluates the efficiency and robustness of our 3D object recognition approach by comparing it with several state-of-the-art approaches.

TABLE 1. Computation time of four recognition approaches on two datasets.

	Bologna dataset				UWA recognition dataset			
Approaches	SHOT+SS	SGC+SS	SGC+ST	SGC+ASV	SHOT+SS	SGC+SS	SGC+ST	SGC+ASV
Description time(s)	0.86	1.09	1.09	1.09	1.82	2.66	2.66	2.66
Feaure match time(s)	0.44	0.46	0.46	0.46	0.77	0.84	0.84	0.84
Selection time(s)	0.01	0.01	2.13	0.72	0.02	0.02	17.28	1.86
Verification time(s)	18.17	14.38	5.21	5.21	54.42	39.76	25.89	12.17
Total(s)	19.48	15.94	8.89	7.48	57.03	43.28	46.67	17.53

1) RECOGNITION EFFICIENCY

To compare recognition efficiency quantitatively, we record the time of each step in the recognition approaches of SHOT+SS, SGC+SS, SGC+ST, and SGC+ASV on the Bologna dataset (1000 feature points on each model and scene) and UWA recognition dataset (3000 feature points on each model and scene), see Table 1. It shows that SGC+ASV is most efficient among the four recognition approaches, which can recognize a scene within several seconds. This is because the ASV selection process (takes around 1-2 seconds) outputs a set of high-quality correspondences that are likely to be correct, and thus reduces lots of time for the model verification stage (saves around 10-40 seconds). Note that ST is computationally expensive since it requires filling an adjacency matrix with a complexity of $O(N^2)$, where N denotes the number of candidate correspondences. In addition, ST is easily trapped into false isometry-maintained clusters under a high level of occlusion, which takes lots of time. Thus, the total recognition time of SGC+ST is even longer than SGC+SS on the UWA recognition dataset.

2) RECOGNITION ROBUSTNESS

We employ the Bologna dataset [11] to evaluate recognition performance with respect to noise and varying point density, and the UWA recognition dataset [10] to evaluate the performance with respect to occlusion.

a: ROBUSTNESS TO NOISE

To evaluate robustness against noise, we add a Gaussian noise with increasing standard deviation of 0.1, 0.2, 0.3, 0.4 and 0.5 pr to each scene and perform recognition with our approach. Recognition rates are reported in Figure 14(a), where the performance of several state-of-the-art approaches on the same dataset are obtained from [3]. The plot shows that our approach successfully recognizes all the models in each scene under different levels of noise. This is achieved only by RoPS [3] and SHOT [11] based approach, while the others are sensitive to noise.

b: ROBUSTNESS TO VARYING POINT DENSITY

To evaluate robustness against varying point density, we downsample each scene to $\frac{1}{2}$, $\frac{1}{4}$ and $\frac{1}{8}$ of its original

point density and perform recognition with our approach. Recognition rates are reported in Figure 14(b). The plot shows that our approach performs the best and obtains 100% recognition rate under all levels of downsampling.

c: ROBUSTNESS TO OCCLUSION

Figure 15(a-e) presents the candidate models, two example scenes and our recognition results with recovered models on the UWA recognition dataset [10]. All the four models in Figure 15(d) are correctly recognized, including the highly occluded Chicken. When the occlusion level increases, a few challenging models may fail to be recognized such as Parasaurolophus in Figure 15(e). Overall, our approach successfully recognizes 183 objects from 50 real scenes consisting of 188 objects, achieving an average recognition rate of 97.3%.

To evaluate robustness against occlusion quantitatively, we classify objects in the scenes into different groups according to their occlusion levels and report recognition rates of our approach and five state-of-the-art approaches. We implement the SHOT based approach using the same recognition framework and record its recognition rate, while the recognition results of other approaches on the same dateset are obtained from [3], [10]. Figure 15(f) shows our approach outperforms all the other approaches except RoPS, under all levels of occlusion. The performance of RoPS based method is slightly better than ours yet it requires models and scenes are represented as triangulated meshes.

IX. CONCLUSION

We have presented a new approach for recognizing 3D objects in cluttered scenes that integrates three novel components: (1) a novel SGC shape descriptor that is robust against occlusion, noise and varying point density; (2) a powerful correspondence compatibility measure that integrates isometry and three angle-preserving components; and (3) a correspondence selection approach that enforces the compatibility constraints based on auxiliary set voting. By this, our recognition approach is not only robust and efficient, but also performs well under challenging recognition scenarios, which cannot be easily achieved by the state-of-the-arts. Quantitative experiments on several publicly available datasets demonstrate the performance of our recognition approach.

REFERENCES

[1] 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.

[2] P. Song and X. Chen, "Pairwise surface registration using local voxelizer," in *Proc. Pacific Graph.*, 2015, pp. 1–6.

[3] 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.

[4] H. Chen and B. Bhanu, "3D free-form object recognition in range images using local surface patches," *Pattern Recognit. Lett.*, vol. 28, no. 10, pp. 1252–1262, Jul. 2007.

[5] A. Aldoma, F. Tombari, L. Di Stefano, and M. Vincze, "A global hypotheses verification method for 3D object recognition," in *Proc. ECCV*, 2012, pp. 511–524.

[6] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2. Oct. 2005, pp. 1482–1489.

[7] E. Rodolá, A. Albarelli, F. Bergamasco, and A. Torsello, "A scale independent selection process for 3D object recognition in cluttered scenes," *Int. J. Comput. Vis.*, vol. 102, no. 1, pp. 129–145, Mar. 2013.

[8] K. Tang, P. Song, and X. Chen, "Signature of geometric centroids for 3d local shape description and partial shape matching," in *Proc. ACCV*, 2016, pp. 1–16.

[9] A. S. Mian, M. Bennamoun, and R. A. Owens, "A novel representation and feature matching algorithm for automatic pairwise registration of range images," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 19–40, Jan. 2006.

[10] A. S. Mian, M. Bennamoun, and R. Owens, "Three-dimensional model-based object recognition and segmentation in cluttered scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1584–1601, Oct. 2006.

[11] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *Proc. ECCV*, 2010, pp. 356–369.

[12] F. Tombari and L. D. Stefano, "Hough voting for 3D object recognition under occlusion and clutter," *IPSJ Trans. Comput. Vis. Appl.*, vol. 4, pp. 20–29, 2012.

[13] J. Glover and S. Popovic, "Bingham procrustean alignment for object detection in clutter," in *Proc. IROS*, Nov. 2013, pp. 2158–2165.

[14] A. E. Johnson, "Spin-Images: A representation for 3-D surface matching," Ph.D. dissertation, Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Aug. 1997.

[15] S. M. Yamany and A. A. Farag, "Surface signatures: An orientation independent free-form surface representation scheme for the purpose of objects registration and matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1105–1120, Aug. 2002.

[16] A. Frome, D. Huber, R. Kolluri, and T. Bülow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *Proc. ECCV*, 2004, pp. 224–237.

[17] 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.

[18] 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.

[19] A. G. Buch, Y. Yang, and N. Krüger, and H. G. Petersen, "In search of inliers: 3D correspondence by local and global voting," in *Proc. CVPR*, Jun. 2014, pp. 2075–2082.

[20] L. Torresani, V. Kolmogorov, and C. Rother, "Feature correspondence via graph matching: Models and global optimization," in *Proc. ECCV*, 2008, pp. 596–609.

[21] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACm Trans. Graph.*, vol. 21, no. 4, pp. 807–832, Oct. 2002.

[22] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3D recognition and pose using the viewpoint feature histogram," in *Proc. IROS*, Oct. 2010, pp. 2155–2162.

[23] C. Papazov and D. Burschka, "An efficient RANSAC for 3D object recognition in noisy and occluded scenes," in *Proc. ACCV*, 2010, pp. 135–148.

[24] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3D object recognition," in *Proc. CVPR*, Jun. 2010, pp. 998–1005.

[25] Y. Guo, F. Sohel, M. Bennamoun, J. Wan, and M. Lu, "A novel local surface feature for 3D object recognition under clutter and occlusion," *Inf. Sci.*, vol. 293, pp. 196–213, Feb. 2015.

[26] 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, no. 2, pp. 348–361, Sep. 2010.

[27] J. Kleinberg, "The small-world phenomenon: An algorithmic perspective," in *Proc. 32nd Annu. ACM Symp. Theory Comput.*, May 2000, pp. 163–170.

[28] Y. Malkov, A. Ponomarenko, A. Logvinov, and V. Krylov, "Approximate nearest neighbor algorithm based on navigable small world graphs," *Inf. Syst.*, vol. 45, pp. 61–68, Sep. 2014.

[29] Y. A. Malkov and D. A. Yashunin. (Mar. 2016). "Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs." [Online]. Available: https://arxiv.org/abs/1603.09320

[30] P. J. Besl and D. N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
[31] *NMSLIB*. [Online]. Available: https://github.com/searchivarius/nmslib
[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] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

KEKE TANG received the B.S. degree in software engineering from Jilin University in 2012. He is currently pursuing the Ph.D. degree with the University of Science and Technology of China, Hefei, China. His research interests include 3-D object recognition, shape matching, robotic vision, and semantic segmentation.

PENG SONG received the B.S. and M.S. degrees from the Harbin Institute of Technology. Shenzhen, in 2007 and 2009, respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2013. He is currently an Associate Professor with the University of Science and Technology of China. His research interests lie in computer graphics, computer vision, and human computer interaction.

XIAOPING CHEN received the B.A. degree in mathematics and the master's degree in electrical engineering from Anhui University, and the Ph.D. degree in computer science from the University of Science and Technology of China (USTC). He established and has led the USTC Multi-Agent Systems Laboratory and the WrightEagle RoboCup Team. He is currently a Professor with the School of Computer Science and Technology, USTC, and also the Director of the Center for Artificial Intelligence Research. He has been a Trustee of RoboCup Federation, since 2008. He has authored about 100 papers. His research interests include intelligent robots and multiagent systems, especially with DEC-POMDPs, NLP, and ASP. He is on the Editorial Board of the Journal of Artificial Intelligence Research and Knowledge Engineering Review.