An Accurate and Robust Range Image Registration Algorithm for 3D Object Modeling

Yulan Guo, Ferdous Sohel, Mohammed Bennamoun, Jianwei Wan, and Min Lu

***Abstract*—Range image registration is a fundamental research topic for 3D object modeling and recognition. In this paper, we propose an accurate and robust algorithm for pairwise and multi-view range image registration. We first extract a set of Rotational Projection Statistics (RoPS) features from a pair of range images, and perform feature matching between them. The two range images are then registered using a transformation estimation method and a variant of the Iterative Closest Point (ICP) algorithm. Based on the pairwise registration algorithm, we propose a shape growing based multi-view registration algorithm. The seed shape is initialized with a selected range image and then sequentially updated by performing pairwise registration between itself and the input range images. All input range images are iteratively registered during the shape growing process. Extensive experiments were conducted to test the performance of our algorithm. The proposed pairwise registration algorithm is accurate, and robust to small overlaps, noise and varying mesh resolutions. The proposed multi-view registration algorithm is also very accurate. Rigorous comparisons with the state-of-the-art show the superiority of our algorithm.**

***Index Terms*—Range image registration, 3D modeling, feature detection, feature description, object reconstruction.**

I. Introduction

THREE dimensional (3D) models are commonly used to describe shapes of objects. This has numerous application areas including education (e.g., electronic museums and multimedia books), entertainment (e.g., 3D TV, games and movies), cultural heritage (e.g., reconstruction of historical relics), medical industry (e.g., orthodontics and diagnosis), manufacturing (e.g., prototyping and inspection) and robotics (e.g., navigation and object/face recognition) [1]–[5]. A model can be built using either Computer Aided Design (CAD) tools or 3D scanning equipments. 3D scanning techniques are the best choice when dealing with free-form objects. However, the range image acquired from a single viewpoint cannot represent the complete shape of an object. Therefore, a 3D object modeling technique is required to register and integrate the set of range images that are acquired from the different viewpoints [6]–[8].

Manuscript received date; revised date. This work was supported by a China Scholarship Council (CSC) scholarship and Australian Research Council grants (DE120102960, DP110102166).

Y. Guo is with the College of Electronic Science and Engineering, National University of Defense Technology, Changsha, Hunan, China, and also with the School of Computer Science and Software Engineering, The University of Western Australia (e-mail: yulan.guo@nudt.edu.cn).

F. Sohel and M. Bennamoun are with the School of Computer Science and Software Engineering, The University of Western Australia (e-mail: ferdous.sohel@uwa.edu.au, mohammed.bennamoun@uwa.edu.au).

J. Wan and M. Lu are with the College of Electronic Science and Engineering, National University of Defense Technology, Changsha, Hunan, China (e-mail: kermitwjw@139.com, lumin@nudt.edu.cn).

Range image registration is a key step for any 3D object modeling system [6]. According to the number of input range images, existing registration algorithms can be classified into pairwise and multi-view registration techniques [9]. Both of them involve two steps: coarse and fine registration [10]. The aim of coarse registration is to estimate an initial transformation between two range images. The resultant initial transformation is then further refined using a fine registration algorithm. Coarse registration can be achieved either manually or automatically [11], [12]. The manual algorithms require human intervention (e.g., a calibrated scanner and turntable, or attached markers) to determine the initial transformation between any two overlapping range images [10]. Their applications are therefore strictly limited due to scenarios where the object must be placed in a fully controlled environment [10], [12]. In contrast, automatic algorithms estimate the initial transformation directly from the data based on the matching of local surface features (also known as correspondence identification [11]). They are more applicable to real-world scenarios compared to their manual counterparts [10]. On that basis, the focus of this paper is on fully automatic range image registration based on local surface features.

A number of local features based pairwise range image registration algorithms have been proposed in the literature [11], [13]–[15]. However, many of these features suffer from low descriptiveness, and/or weak robustness to certain nuisances including noise and varying mesh resolutions [15], [16] (see Section II). Besides, lots of multi-view range image registration algorithms can also be found in the literature, e.g., the spanning tree based algorithm [9], [15], [17], [18]. One major limitation of these algorithms is their high computational complexity due to the expensive exhaustive search [11]. Mian et al. proposed a connected graph and hypergraph based algorithm [11], which is more computationally efficient compared to [9]. However, it can only be applied to the cases when the given range images are from a single object or scene.

In this paper, we propose a fully automatic, accurate and robust (pairwise and multi-view) range image registration algorithm for the simultaneous modeling of multiple 3D objects. That is, given a set of mixed and unordered range images (where each range image includes only one object), our algorithm automatically registers all the range images which are related to the same object. This paper first uses a feature called Rotational Projection Statistics (RoPS) [19] for pairwise range image registration algorithm. It exhibits both high accuracy and strong robustness to noise and varying mesh resolutions. A comparison with a set of the state-of-the-art algorithms shows the superiority of our algorithm. Based on the pairwise registration algorithm, a shape growing

based fully automatic multi-view range image registration algorithm is proposed. Extensive experiments were performed to demonstrate the effectiveness and efficiency of the multi-view registration algorithm.

The rest of this paper is organized as follows. Section II briefly reviews related work on range image registration. Section III introduces a RoPS based pairwise range image registration algorithm. Section IV describes a shape growing based multi-view range image registration algorithm. Section V presents the experimental results of our proposed algorithms, with comparison to the state-of-the-art.

II. Related Work

This section presents a brief overview of the existing algorithms for pairwise and multi-view range image registration.

A. Pairwise Registration Algorithms

1) Coarse Registration: Fully automatic pairwise coarse registration is usually accomplished by finding point correspondences through the matching of local features [20]. A number of local surface features have been introduced in the literature. Stein and Medioni [21] proposed "splash" by describing the distribution of surface normals along a geodesic circle. Johnson and Hebert [13] represented the neighboring points of a feature point with a cylindrical coordinate frame, and proposed a "spin image" feature to encode the local surface. Spin image is among the most cited local feature extraction algorithms. However, it suffers from several limitations including its low descriptiveness and high sensitivity to varying mesh resolutions [19]. Following a similar approach to [13], Yamany and Farag [14] proposed "surface signatures" to encode a local surface. Frome et al. [22] introduced a "3D Shape Context (3DSC)" by counting up the weighted number of neighboring points falling into the bins of a spherical space. One major limitation of 3DSC is its uncertainty in the rotation around the surface normal [23]. Later, Tombari et al. [23] improved 3DSC by constructing a unique reference frame for each feature, resulting in a "Unique Shape Context (USC)" feature. Chen and Bhanu [24] proposed "Local Surface Patches (LSP)" to represent the shape index values and normal variations in the local surface. Rusu et al. [25] introduced "Point Feature Histograms (PFH)" by encoding the relative information (i.e., angles and a distance) between all pairs of the neighboring points. They then proposed a FPFH feature to improve the computational efficiency of PFH [26]. Tombari et al. [18] divided the neighborhood space of a feature point into 3D spherical volumes, and used the angles between the normal of the feature point and these of the neighboring points to generate a SHOT feature descriptor. However, many of the existing features suffer from either low matching accuracy, or high sensitivity to certain nuisances including noise and varying mesh resolutions [15], [19]. For more details on 3D local surface features, the reader is referred to a comprehensive and contemporary survey [27].

2) Fine Registration: Once an estimation of the transformation between two range images is obtained, a fine pairwise registration algorithm is performed to produce a more accurate solution. Several direct solutions (which did not require any iterative calculations) were proposed in the literature [28]. However, the most popular algorithms use iterative approaches to achieve more accurate registration results, e.g., [29], [30]. Besl and McKay [30] proposed an ICP algorithm to minimize the average point-to-point distance of the closest point pairs between two range images. However, the original ICP algorithm requires range images to have a significant overlap, and is not robust to outliers [31]. Later, a set of variants have been proposed to improve the performance of the original ICP algorithm [32]. Chen and Medioni [29] proposed a fine registration algorithm based on the minimization of the average point-to-plane distance rather than the point-to-point distance between two range images. Compared to the ICP algorithm, Chen and Medioni's algorithm usually requires less iterations to reach convergence [10]. However, solving the nonlinear least squares problem to minimize the residual registration error is computationally expensive [31]. Liu [33] proposed a fine registration algorithm based on the Lyapunov function for a Markov chain of thermodynamic systems. A comparative study shows that Liu's algorithm outperforms the state-of-the-art ICP variants. However, it is also the most computationally expensive algorithm compared to the state-of-the-art. Therefore, we use a variant of the ICP algorithm for the fine registration.

B. Multi-view Registration Algorithms

1) Coarse Registration: A multi-view coarse registration algorithm involves two tasks. The first task is to recover the overlap information between the input range images, and the second task is to calculate the rigid transformations between any two overlapping range images (which is effectively a pairwise registration). Huber and Hebert [9] first applied the spin image based pairwise coarse registration algorithm to all pairs of range images to construct a model graph. They then searched this graph for a spanning tree which was pose consistent and globally surface consistent. This spanning tree was finally used to register multi-view range images. Following the same approach, Masuda [17], Bariya et al. [15] and Tombari et al. [18] respectively used LPHM, SHOT and exponential map based algorithms for pairwise coarse registration, and constructed a spanning tree of the input range images that maximize the sum of the number of inlier point pairs (or the area of overlap). For a set of N_m range images, the computational complexity of the Huber and Hebert's spanning tree based algorithm and its variants is $O\left(N_m^2\right)$ as they need to exhaustively register every pair of range images. These algorithms are therefore very time consuming and infeasible for the registration of a large number of range images [9]. Mian et al. [11] constructed a connected graph by choosing the range image with the maximum surface area as the root node and iteratively added new range images with enough corresponding 3D tensors and passed global verification to the graph. Guo et al. [34] followed a similar technique and used Tri-Spin-Image features to perform multi-view registration. ter Haar [35] selected quadruples of range images to form incomplete 3D models of an object. These quadruples were then verified

and aligned to obtain the final alignment. This algorithm was more efficient compared to [9]. One major limitation of this algorithm is that each quadruple should cover the entire object, and range images which cover a small part of an object cannot be registered [35].

2) Fine Registration: Based on the multi-view coarse registration results, multi-view fine registration algorithms aim to minimize the registration error of all overlapping range images. Benjemaa and Schmitt [36] extended the ICP algorithm, Neugebauer [37] extended the Chen and Medioni's algorithm from pairwise fine registration to multi-view fine registration. Williams and Bennamoun [38] proposed an extension of the Arun et al.'s pairwise registration algorithm [28] to perform simultaneous registration of multiple corresponding point sets. Masuda [39] represented the input range images with Signed Distance Field (SDF) samples, and simultaneously performed multi-view registration and integration based on the SDF samples. Nishino and Ikeuchi [40] employed the laser reflectance strength as an additional attribute of the 3D points to further improve the robustness of a multi-view registration algorithm.

C. Major Contributions

This paper has three major contributions, which are summarized as follows.

(i) We introduce an accurate pairwise range image registration algorithm based on RoPS features. The algorithm is very robust to small overlaps, noise and variations in the mesh resolutions. Comparisons with a set of state-of-the-art algorithms prove the superiority of our image registration algorithm. Although RoPS [19] was originally proposed for 3D object recognition, we use it in this paper for 3D object modeling without any prior information about the scene (e.g., order of the range images). The effectiveness and robustness of the RoPS descriptor has been fully demonstrated in the context of 3D modeling (through extensive experimental tests).

(ii) We propose a multi-view range image registration algorithm based on shape growing (Section IV). Although the concept of shape growing has been used in other work (e.g., [6]) for the registration of successive range images of a single object, we propose in contrast a novel multi-view registration algorithm for the registration of a set of mixed and unordered range images from multiple 3D objects.

(iii) We integrate our pairwise and multi-view range image registration algorithms into a framework for automatic 3D object modeling (as shown in Fig. 1). The framework was tested on a number of range images acquired with both high and low-resolution sensors (i.e., Cyberware, Minolta Vivid, Microsoft Kinect, and Space Time). Experimental results show that the proposed framework is able to reconstruct the 3D model of an object without any manual intervention (Section V-C for more details).

III. Pairwise Range Image Registration

A pairwise registration algorithm should be automatic and accurate. It also should be robust to small overlaps, noise, varying mesh resolutions and other nuisances. In this section, we introduce a RoPS based pairwise registration algorithm which satisfies these qualifications (see Section V-A). The algorithm consists of four parts: RoPS feature extraction, feature matching, robust transformation estimation and fine registration. The reader is referred to [19] for more details on the RoPS feature extraction and feature matching. It is also briefly described below for completeness.

A. RoPS Feature Extraction

Given a range image $\mathcal{I}_i$ or a pointcloud generated from this, it has to be converted into a triangular mesh $\mathcal{M}_i$ since the subsequent feature point detection and feature description algorithms work on mesh data. This can be achieved by Delaunay triangulation [19] or the Marching Cubes algorithm [41]. We then detect a set of feature points $\boldsymbol{p}_k^i, k = 1, 2, \ldots, N_i$ from $\mathcal{M}_i$ and represent these points using our previously proposed RoPS feature descriptors [19], [42].

In order to detect unique and repeatable feature points, the mesh $\mathcal{M}_i$ is first simplified to a low-resolution mesh $\widehat{\mathcal{M}}_i$, the vertices of $\mathcal{M}_i$ which are nearest to the vertices of $\widehat{\mathcal{M}}_i$ are selected as candidate points. These candidate points are then filtered by a resolution control technique [19] to remove redundant points. Boundary points are also pruned out from these candidate points to improve their stability. For each remaining point, we perform a weighted Continuous Principal Component Analysis (CPCA) [19] on its local surface, resulting in three eigenvalues λ_1, λ_2 and λ_3. The points with $\frac{\lambda_1}{\lambda_2} > \tau_\lambda$ are considered as feature points, where τ_λ is a threshold to further remove the points with symmetrical underlying local surfaces. The threshold τ_λ determines both the number and the stability of the feature points. That is, a large threshold can produce a limited number of feature points with high stability. In this papers τ_λ is empirically set to 1.02. For each feature point $\boldsymbol{p}_k^i$ in mesh $\mathcal{M}_i$, a local surface $\mathcal{L}_k^i$ is first cropped from $\mathcal{M}_i$ for a given support radius r. Then, a unique and unambiguous Local Reference Frame (LRF) $\mathbf{F}_k^i$ is derived using the eigenvectors of its local surface $\mathcal{L}_k^i$. The points on $\mathcal{L}_k^i$ are aligned with this LRF $\mathbf{F}_k^i$ to make the resultant feature descriptor invariant to rotation and translation. The transformed local surface $\widetilde{\mathcal{L}}_k^i$ are then used to construct a RoPS feature descriptor.

The local surface $\widetilde{\mathcal{L}}_k^i$ is first rotated around the x axis by a set of angles. For each rotation, the points on $\widetilde{\mathcal{L}}_k^i$ are projected onto three coordinate planes (i.e., the xy, xz and yz planes) to extract several statistics. Specifically, we first obtain an $L \times L$ distribution matrix $\mathbf{D}$ of the projected points on each plane, and then calculate five statistics (including central moments μ_{11}, μ_{21}, μ_{12}, μ_2 and entropy e) for the distribution matrix $\mathbf{D}$. That is,

$$\mu_{mn} = \sum_{a=1}^{L}\sum_{b=1}^{L} (a-\bar{a})^m \left(b-\bar{b}\right)^n \mathbf{D}\left(a,b\right), \tag{1}$$

$$e = -\sum_{a=1}^{L}\sum_{b=1}^{L} \mathbf{D}\left(a,b\right) \log\left(\mathbf{D}\left(a,b\right)\right). \tag{2}$$

where

$$\bar{a} = \sum_{a=1}^{L}\sum_{b=1}^{L} a\mathbf{D}\left(a,b\right), \tag{3}$$

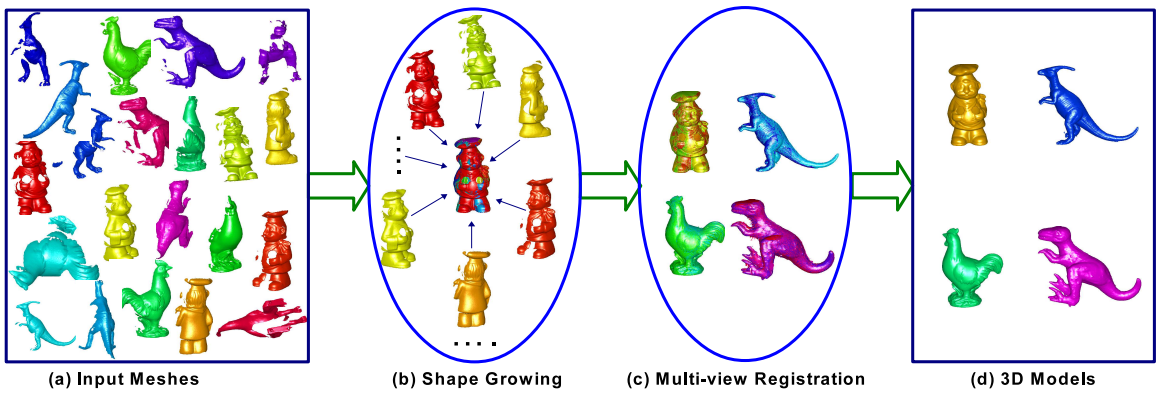


Fig. 1: An illustration of the proposed 3D object modeling framework (Figure best seen in color).

$$\bar{b} = \sum_{a=1}^{L}\sum_{b=1}^{L} b\mathbf{D}\left(a,b\right). \tag{4}$$

These statistics for all coordinate planes and rotations are concatenated to form a sub-feature $\boldsymbol{f}_{kx}^{i}$. The local surface $\widetilde{\mathcal{L}}_{k}^{i}$ is then rotated around the y and z axes respectively to obtain another two sub-features $\boldsymbol{f}_{ky}^{i}$ and $\boldsymbol{f}_{kz}^{i}$. The overall RoPS feature $\boldsymbol{f}_{k}^{i}$ is obtained by concatenating the three sub-features $\boldsymbol{f}_{kx}^{i}$, $\boldsymbol{f}_{ky}^{i}$ and $\boldsymbol{f}_{kz}^{i}$. The reader is referred to [19] for further details about the RoPS feature extraction.

B. Feature Matching

Let $\boldsymbol{F}^{i} = \left\{\boldsymbol{f}_{1}^{i}, \boldsymbol{f}_{2}^{i}, \ldots, \boldsymbol{f}_{N_i}^{i}\right\}$ and $\boldsymbol{F}^{j} = \left\{\boldsymbol{f}_{1}^{j}, \boldsymbol{f}_{2}^{j}, \ldots, \boldsymbol{f}_{N_j}^{j}\right\}$ respectively be two sets of RoPS features for meshes $\mathcal{M}_i$ and $\mathcal{M}_j$. For a feature $\boldsymbol{f}_{k}^{i}$ from $\boldsymbol{F}^{i}$, we find its nearest feature $\boldsymbol{f}_{k}^{j}$ from $\boldsymbol{F}^{j}$ as:

$$\boldsymbol{f}_{k}^{j} = \arg\min_{n=1,2,\ldots,N_j} \left(\left\|\boldsymbol{f}_{k}^{i} - \boldsymbol{f}_{n}^{j}\right\|_{2}\right). \tag{5}$$

The pair $\left(\boldsymbol{f}_{k}^{i}, \boldsymbol{f}_{k}^{j}\right)$ are considered a feature correspondence, and their associated points $\boldsymbol{c}_{k}^{ij} = \left(\boldsymbol{p}_{k}^{i}, \boldsymbol{p}_{k}^{j}\right)$ are considered a point correspondence. Note that, there might be more than one nearest feature in $\boldsymbol{F}^{j}$ for a given $\boldsymbol{f}_{k}^{i}$. In that case, several point correspondences can be generated for the feature $\boldsymbol{f}_{k}^{i}$. We employed the k-d tree algorithm to reduce the computational complexity of feature matching. All features in $\boldsymbol{F}^{i}$ are matched against these features in $\boldsymbol{F}^{j}$, resulting in a set of point correspondences $\mathcal{C}^{ij} = \left\{\boldsymbol{c}_{1}^{ij}, \boldsymbol{c}_{2}^{ij}, \ldots, \boldsymbol{c}_{N_i}^{ij}\right\}$.

For each point correspondence $\boldsymbol{c}_{k}^{ij}$, a rigid transformation $\mathbf{T}_{k}^{ij} = \left(\mathbf{R}_{k}^{ij}, \boldsymbol{t}_{k}^{ij}\right)$ can be calculated using their point positions $\left(\boldsymbol{p}_{k}^{i}, \boldsymbol{p}_{k}^{j}\right)$ and LRFs $\left(\mathbf{F}_{k}^{i}, \mathbf{F}_{k}^{j}\right)$. That is,

$$\mathbf{R}_{k}^{ij} = \left(\mathbf{F}_{k}^{i}\right)^{\mathrm{T}} \mathbf{F}_{k}^{j}, \tag{6}$$

$$\boldsymbol{t}_{k}^{ij} = \boldsymbol{p}_{k}^{i} - \boldsymbol{p}_{k}^{j}\mathbf{R}_{k}^{ij}, \tag{7}$$

where $\mathbf{R}_{k}^{ij}$ is the rotation matrix and $\boldsymbol{t}_{k}^{ij}$ is the translation vector of the rigid transformation $\mathbf{T}_{k}^{ij}$. Totally, N_i transformations can be calculated from these point correspondences. Note that, several incorrect point correspondences may exist in $\mathcal{C}^{ij}$, which will subsequently result in wrong transformation estimations.

C. Robust Transformation Estimation

Several methods have been developed in the literature to produce a correct transformation from a set of point correspondences with outliers, e.g., the Random Sample Consensus (RANSAC) method [43] and its variants [44], and the rigidity constraint based method [8]. We instead use a Consistent Correspondences Verification (CCV) method [19], which is shown to perform better than the RANSAC method (see Section V-A1).

Let $\mathcal{C}^{ij} = \left\{\boldsymbol{c}_{1}^{ij}, \boldsymbol{c}_{2}^{ij}, \ldots, \boldsymbol{c}_{N_i}^{ij}\right\}$ be the point correspondence set for the mesh pair $\mathcal{M}_i$ and $\mathcal{M}_j$, and $\mathbf{T}_{k}^{ij}$ be the estimated transformation from point correspondence $\boldsymbol{c}_{k}^{ij}$. For each estimated transformation $\mathbf{T}_{k}^{ij}$, we find out all point correspondences $\mathfrak{C}_{k}^{ij}$ whose estimated transformations are similar to $\mathbf{T}_{k}^{ij}$. Specifically, we first convert the rotation matrix of each transformation into three Euler angles. We then measure the difference between any two transformations using both the distance d_a between their Euler angles and the distance d_t between their translation vectors. The transformations, whose angle distances d_a to $\mathbf{T}_{k}^{ij}$ are less than a threshold τ_a and translation distances d_t to $\mathbf{T}_{k}^{ij}$ are less than a threshold τ_t, are selected to form a group of consistent correspondences $\mathfrak{C}_{k}^{ij}$. In this paper, τ_a and τ_t are empirically set to 0.2 and $10d_{res}$, respectively. We then calculate a plausible transformation estimation $\widetilde{\mathbf{T}}_{k}^{ij}$ for each point correspondence $\boldsymbol{c}_{k}^{ij}$ using the group of consistent correspondences $\mathfrak{C}_{k}^{ij}$. If $\mathfrak{C}_{k}^{ij}$ has at least three point correspondences, the $\widetilde{\mathbf{T}}_{k}^{ij}$ is calculated from these point correspondences using a least-square fitting method [45]. Otherwise, $\widetilde{\mathbf{T}}_{k}^{ij}$ is calculated as the average value of these transformation estimations corresponding to $\mathfrak{C}_{k}^{ij}$.

In order to find out the best transformation estimation from $\mathfrak{C}^{ij} = \{\mathfrak{C}_{1}^{ij}, \mathfrak{C}_{2}^{ij}, \ldots, \mathfrak{C}_{N_i}^{ij}\}$, $\mathcal{M}_i$ and $\mathcal{M}_j$ are first simplified

to two low-resolution meshes $\widetilde{\mathcal{M}}_i$ and $\widetilde{\mathcal{M}}_j$, respectively. The process of mesh simplification is performed using the MATLAB function "*reducepatch*", which reduces the number of points and preserves the overall shape of the mesh. Each transformation $\widetilde{\mathbf{T}}_k^{ij}$ is then used to align mesh $\widetilde{\mathcal{M}}_i$ to $\widetilde{\mathcal{M}}_j$. The one which results in the maximum number of inlier point pairs is considered the final transformation estimation $\mathbb{T}^{ij}$. Here, a pair of points is defined as an inlier point pair only if the distance between the two points is less than two times the average mesh resolution. In order to find inlier point pairs, we match only the points in the simplified meshes ($\widetilde{\mathcal{M}}_i$ and $\widetilde{\mathcal{M}}_j$) for computational efficiency. Since the numbers of vertices in the simplified meshes are much smaller compared to their original meshes ($\mathcal{M}_i$ and $\mathcal{M}_j$), this process of simplification greatly reduces the computational time for alignment.

D. Fine Registration

Once the initial transformation $\mathbb{T}^{ij}$ is determined, a variant of the ICP algorithm is used to perform fine registration between the meshes $\mathcal{M}_i$ and $\mathcal{M}_j$. Starting with the initial transformation $\mathbb{T}^{ij}$, the ICP algorithm iteratively refines the rigid transformation by repeatedly generating pairs of closest points in the two meshes and minimizing a residual error (i.e., the average distance between pairs of closest points) [2], [30]. This variant differs from the original ICP algorithm in several aspects. First, we adopt a coarse-to-fine sampling approach to improve its computational efficiency. Rather than using all points in $\mathcal{M}_i$ to search for their closest points in $\mathcal{M}_j$, we take only a part of the points from $\mathcal{M}_i$ at the m-th iteration. Since random subsampling and uniform subsampling based ICP algorithms have a very similar registration performance [32], we only use random sampling for its simplicity.

The number of sample points n_m is defined as:

$$n_m = -\frac{1}{2}\log\left(\frac{\epsilon_{m-1}}{d_{res}}\right)(n_{max} - n_{min}) + n_{min}, \quad (8)$$

where n_{max} and n_{min} are respectively the pre-defined maximum and minimum number of sampling points for each iteration. ϵ_{m-1} is the residual error for the last iteration, which is usually smaller than the average mesh resolution d_{res}. According to Eq. (8), the number of sample points in each iteration is related to the residual error. That is, we use initially a small number of sample points in order to speedup the process when the residual error is still very large. Then, when the residual error decreases, we use more sample points to improve the registration accuracy. In order to further improve the accuracy and stability of the ICP algorithm, we reject at each iteration all pairs of closest points whose point-to-point distances are more than 2 times the average mesh resolution.We also use the k-d tree algorithm to perform a closest point search efficiently.

IV. Multi-view Range Image Registration

So far we have described our proposed RoPS based pairwise registration algorithm. This registration algorithm is used as the basis for our proposed multi-view range image registration algorithm. The process is described illustrated in Fig. 1.

A. Shape Growing based Coarse Registration

Given a set of input meshes $\{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_{N_m}\}$, the task of multi-view range image registration is to register them to a common coordinate frame effectively. The algorithm starts by initializing the search space $\mathbf{\Phi}$ with all the input meshes, and then selects a mesh from the search space as the seed shape $\boldsymbol{\mathcal{R}}_1$. The shape $\boldsymbol{\mathcal{R}}_1$ iteratively grows by performing pairwise registration between itself and the remaining meshes in the search space.

For a mesh $\mathcal{M}_i$ in the search space, we use the RoPS based pairwise registration algorithm (which includes the ICP fine alignment) to register it to the shape $\boldsymbol{\mathcal{R}}_1$ (as shown in Fig. 2(a)). The mesh $\mathcal{M}_i$ is considered to be successfully registered to $\boldsymbol{\mathcal{R}}_1$ only if the number of overlapping points exceeds a predefined threshold (e.g., 0.5 times of the number of vertices in $\mathcal{M}_i$). If the registration is successful, the mesh vertices in $\mathcal{M}_i$ (as shown in Fig. 2(b)), whose shortest distances to the registered $\boldsymbol{\mathcal{R}}_1$ are larger than the average mesh resolution, are added to the shape $\boldsymbol{\mathcal{R}}_1$. Consequently, the shape $\boldsymbol{\mathcal{R}}_1$ is updated. The updated shape contains all points of the previous shape and some points of the input mesh $\mathcal{M}_i$. Note that, the distance constraint used here is to make sure that no redundant point is added to the shape.

We now need to extract RoPS features for the newly updated shape $\boldsymbol{\mathcal{R}}_1$. Since the RoPS features of all these input meshes have already been extracted, we therefore, generate RoPS features for $\boldsymbol{\mathcal{R}}_1$ from the already available RoPS features of the previous shape and the mesh $\mathcal{M}_i$ (rather than following the method described in Section III-A to extract brand-new features). Specifically, for each feature point in the previous shape and the mesh $\mathcal{M}_i$, we find its closest point (corresponding point) in the updated shape. If the distance between the two points is less than the average mesh resolution, the corresponding point in the updated shape is considered a feature point. Note that, the distance constraint used here is to make sure that the two points correspond to the same physical position. We also applied a resolution control strategy [19] to the newly added feature points to remove any redundant feature points. Once the feature point in the updated shape is selected, the LRF and RoPS feature descriptor of the existing feature point in the previous shape or the mesh $\mathcal{M}_i$ are assigned to the corresponding feature point in the updated shape (as shown in Fig. 2(b) and (c)). Note that, this feature extraction process for the shape $\boldsymbol{\mathcal{R}}_1$ can be performed immediately once the shape is updated. Therefore, this process improves the computational efficiency of feature extraction as it does not need any feature calculation (as described in Section III-A) during the process of shape growing.

Once the mesh $\mathcal{M}_i$ is checked, it is then removed from the search space $\mathbf{\Phi}$. If the registration is successful, the transformation information between $\mathcal{M}_i$ and the shape $\boldsymbol{\mathcal{R}}_1$ is stored. The algorithm then proceeds to the next unchecked mesh $\mathcal{M}_{i+1}$ in the search space $\mathbf{\Phi}$, and the shape growing process is performed again on the newly selected mesh $\mathcal{M}_{i+1}$. This iterative process of shape growing continues until either all the meshes have been registered to $\boldsymbol{\mathcal{R}}_1$, or no mesh in the search space $\mathbf{\Phi}$ can further be registered to $\boldsymbol{\mathcal{R}}_1$. It is

(a) (b) (c)

Fig. 2: An illustration of the process of shape growing. (a) The seed shape. (b) The input mesh, where the red dots represent the points which are going to be updated to the seed shape. (c) The updated shape. The blue dots represent the corresponding points between the input mesh and the updated shape (Figure best seen in color).

worth noting that, the surface of the shape $\mathcal{R}_1$ gradually grows into a final 3D shape (rather than 2.5D), as shown in Fig. 1(b). Meanwhile, the pose of the shape $\mathcal{R}_1$ keeps unchanged during the process. Therefore, all range images are registered to a common coordinate frame (i.e., the one used by $\mathcal{R}_1$). Once the process of shape growing stops, the rigid transformations between all these registered meshes and $\mathcal{R}_1$ are already known. We then transform these meshes to the coordinate frame of $\mathcal{R}_1$. Consequently, these meshes are coarsely registered. Note that, the shape $\mathcal{R}_1$ is just used as a reference (i.e., a common coordinate frame), it will not be used for the subsequent stages (e.g., fine registration and 3D modeling). We just use the coarsely registered meshes for the 3D modeling.

In order to cope with the cases where the meshes may correspond to several different objects, the algorithm continues to initialize a new seed shape $\mathcal{R}_2$ by picking up a mesh from the remaining meshes in the search space. The shape $\mathcal{R}_2$ grows using the same technique as for $\mathcal{R}_1$. Consequently, all the meshes corresponding to the shape $\mathcal{R}_2$ are coarsely registered. This process continues until no initial seed shape can be built anymore. Finally, all these input meshes can separately be registered to their corresponding shapes. That is, the meshes which correspond to a particular shape are considered to be from the same object.

Compared to the spanning tree based algorithms (e.g., [9], [15], [17], [18]), the advantages of the shape growing based algorithm are at least twofold. First, it performs coarse registration on range images more efficiently, as demonstrated in Section V-B3. Second, it can accomplish multi-view registration of range images corresponding to several different objects, rather than from only a single object, as further demonstrated in Section V-B4.

Other related work includes [6], [3] and [7]. Our algorithm differs from these methods in several aspects. First, [6], [3] and [7] mainly focus on the registration of successive range images. In contrast, our multi-view registration algorithm can successfully register a set of unordered range images. Second, [6] and [3] work on the registration of range images from a single object, while our algorithm can work on a set of mixed range images corresponding to multiple objects (as demonstrated in V-B4). Third, [6], [3] and [7] reconstructed a single 3D model/scene in the context of user interaction. In contrast, this paper focuses on offline automatic (non-interactive) modeling of multiple 3D objects. Fourth, [6], [3] and [7] directly used the updated shape (surface) for the final reconstructed model. We however, use the updated shape as the reference of a common coordinate system for multi-view coarse registration, and then perform 3D reconstruction from the transformed input meshes. Therefore, we can avoid the accumulated error caused by shape growing. Besides, [7] worked on depth images which have regular lattices while our algorithm works on pointclouds (or meshes).

B. Fine Registration and 3D Modeling

Once the meshes corresponding to a particular shape are coarsely registered, these registrations are refined with a multi-view fine registration algorithm (e.g., [38]). This process further minimizes the overall registration error of multiple meshes, and distributes any registration errors evenly over the complete 3D model. A continuous and seamless 3D model is finally reconstructed for each shape by using an integration and surface reconstruction algorithm [46].

Note that, the proposed algorithm is fully automatic and can be performed without any manual intervention. It does not require any prior information about the sensor position, the shapes of objects, viewing angles, overlapping pairs, order of meshes, or number of objects. In our case, a user can treat the modeling process as a "black box". The only thing one needs to do is to import all scanned range images to the system, and to collect the complete 3D models at the output.

V. Experimental Results And Analysis

In this section, a set of experiments were performed to test the effectiveness and robustness of our proposed algorithms.

A. Pairwise Registration Results

We tested the performance of our pairwise registration algorithm on the UWA 3D Modeling Dataset [11]. The dataset consists of 22, 16, 16, and 21 range images respectively from four objects, namely the Chef, Chicken, Parasaurolophus and T-Rex. These range images were acquired with a Minolta Vivid 910 scanner. In order to calculate the ground truth rotation $\mathbf{R}_{GT}^{ij}$ and translation $\boldsymbol{t}_{GT}^{ij}$ between any two range images $\mathcal{M}_i$ and $\mathcal{M}_j$, the two range images were first aligned manually and then further refined using the ICP algorithm. Their degree of overlap was calculated as the ratio of overlapping points to the average number of points of the two aligned range images. Note that, we used manual alignment only for the ground truth generation. During the performance tests, all algorithms were conducted automatically without any manual intervention (i.e., the whole process is fully automatic).

We measured the accuracy of a pairwise registration using two errors [16], i.e., the error ϵ_r^{ij} between the estimated rotation $\mathbf{R}_E^{ij}$ and the ground truth rotation $\mathbf{R}_{GT}^{ij}$, and the error

ϵ_t^{ij} between the estimated translation $\boldsymbol{t}_E^{ij}$ and the ground truth translation $\boldsymbol{t}_{GT}^{ij}$. The rotation error ϵ_r^{ij} is calculated as:

$$\epsilon_r^{ij} = \arccos\left(\frac{\text{trace}\left(\mathbf{R}_d^{ij}\right) - 1}{2}\right)\frac{180}{\pi}, \tag{9}$$

where

$$\mathbf{R}_d^{ij} = \mathbf{R}_{GT}^{ij}\left(\mathbf{R}_E^{ij}\right)^{-1}. \tag{10}$$

The translation error ϵ_t^{ij} is calculated as:

$$\epsilon_t^{ij} = \frac{\left\|\boldsymbol{t}_{GT}^{ij} - \boldsymbol{t}_E^{ij}\right\|}{d_{res}}. \tag{11}$$

where d_{res} is the average mesh resolution of all range images of an object.

We performed pairwise registration between any two range images of an object, and calculated their registration errors. The range image pairs with an extremely small overlap (less than 10%) were excluded from the results, the same as in [47]. In total, 145, 83, 82 and 147 pairs of range images respectively from the Chef, Chicken, Parasaurolophus and T-Rex were used in the experiments. A registration was reported as correct if the rotation error was less than 5° and the translation error was less than $5d_{res}$ (except in Section V-A1 where different thresholds are used for the experiments). Otherwise, it was considered as an incorrect registration.

1) Transformation Estimation: We performed pairwise range image registration using different transformation estimation methods including the CCV method and the RANSAC method [43]. During each iteration of the RANSAC method, three pairs of point correspondences are randomly selected to calculate a transformation. The number of inlier point pairs between two simplified meshes is selected as a measure for alignment, the transformation which results in the maximum number of inlier point pairs is selected and then refined using the inlier point pairs to obtain the final transformation. In this paper, we tested the RANSAC method with different numbers of iterations (i.e., 500, 1000, 5000 and 10000). We also tested a variant of the RANSAC method [17] using the same parameters as in the article [17]. In addition, we tested a method which calculated a transformation directly from the whole set of point correspondences using a Least Square Fitting (LSF) approach. An illustration of the pairwise range image registration is shown in Fig. 3.

We report in Table I the percentages of correct registrations of these methods with different thresholds. That is, we considered a registration as correct if the translation error was less than $1d_{res}$, $2d_{res}$, $3d_{res}$, $4d_{res}$, $5d_{res}$, $10d_{res}$, $15d_{res}$, and $20d_{res}$, respectively. Several observations can clearly be made from these results. First, the RANSAC method achieved a significant improvement compared to the LSF method. Taking the results under a threshold of $5d_{res}$ as an example, 45.73% correct registrations were achived by the RANSAC method with 5000 iterations, while only 3.94% correct registrations were achieved by the LSF method. That is because the RANSAC method adopted a consensus check technique to detect and reject outliers, which therefore, greatly improved the accuracy of the transformation estimation. Second, the CCV method outperformed the RANSAC and LSF methods by a large margin under all levels of thresholds. For example, with a strict threshold of $1d_{res}$, 39.17% of the range image pairs can be correctly registered by the CCV method, while only 26.70% can be registered by the RANSAC method with 5000 iterations. As the threshold increased to $5d_{res}$, 67.61% of the range image pairs were correctly registered by the CCV method, while 45.73% were registered by the RANSAC method with 5000 iterations. In the rest of the paper, we used a threshold of $5d_{res}$ to ensure that the RANSAC method has a relatively high percentage of correct registrations. Third, the results achieved by the RANSAC method improved as the number of iterations increased. For example, under a threshold of $5d_{res}$, 32.82% and 47.92% correct registrations were achieved by the RANSAC method with 500 iterations and 10000 iterations, respectively. Fourth, the results achieved by the RANSAC method with 10000 iterations were similar to the results achieved by the variant proposed in [17]. Both of them were inferior to those achieved by our CCV method.

Note that, with the traditional RANSAC method, three point correspondences are randomly selected to derive a plausible transformation. For the CCV method, a group of (usually more than three) consistent correspondences are used to derive a plausible transformation. The plausible transformation which results in the maximum number of inlier point pairs is selected. This transformation is then refined to get the final transformation estimate. Since the CCV method uses more than three consistent point correspondences compared to three randomly selected points for the RANSAC method, the CCV method produces more accurate results compared to the traditional RANSAC method.

(a) (b)

Fig. 3: An illustration of pairwise range image registration. (a) A pair of range images with the correct point correspondences. (b) Registered range images (Figure best seen in color).

2) Comparison with Other Features: We performed pairwise range image registration using different features including the spin image, LSP, THRIFT, USC, FPFH and our RoPS. The registration results are shown in Table II. Our RoPS based algorithm achieved the best registration performance. It produced the largest percentage of correct registrations. The THRIFT based algorithm obtained the second best results in terms of the percentage of correct registrations, followed by the LSP, USC, FPFH and spin image based algorithms. The rotation and translation errors achieved by all these features

TABLE I: Percentage of correct registrations using different transformation estimation methods.

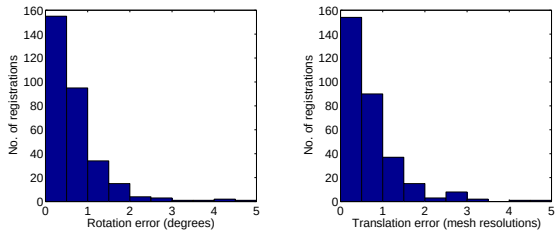
	$1d_{res}$	$2d_{res}$	$3d_{res}$	$4d_{res}$	$5d_{res}$	$10d_{res}$	$15d_{res}$	$20d_{res}$
CCV	39.17	57.99	63.90	65.86	67.61	68.05	68.05	68.05
LSF	1.09	1.53	2.63	3.06	3.94	5.03	6.13	7.22
RANSAC-500	20.57	29.76	31.73	31.95	32.82	35.01	35.67	36.54
RANSAC-1000	22.32	32.39	35.23	37.42	37.64	39.17	39.61	40.04
RANSAC-5000	26.70	37.86	45.67	44.86	45.73	46.83	48.36	48.36
RANSAC-10000	29.76	39.61	45.08	46.17	47.92	49.23	49.89	50.11
RANSAC [17]	29.32	40.92	45.30	46.17	46.61	49.02	50.11	50.33

were comparable. Note that, the superior performance of our RoPS based algorithm is partly due to the high descriptiveness of our RoPS feature descriptors. Since the RoPS feature descriptor encodes more distinctive information of the local surface (as demonstrated in our previous paper [19]) compared to existing feature descriptors, it therefore, produces more reliable and robust feature matching results. Consequently, the registration performance is improved.

TABLE II: Pairwise registration results using different features.

	%registration	Error ϵ_r ($^\circ$)	Error ϵ_t (d_{res})
RoPS	67.61	0.6655	0.6756
Spin image	51.42	0.7562	0.6312
LSP	60.18	0.8944	0.7803
THRIFT	61.93	0.7541	0.6764
USC	58.21	0.7619	0.7046
FPFH	57.77	0.8592	0.7136

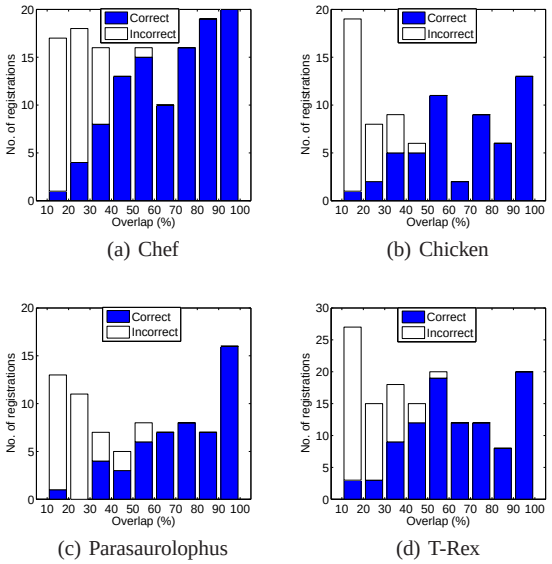
In order to further illustrate the performance of the RoPS based algorithm, we present the histograms of the rotation and transformation errors of the four objects in Fig 4. It can be seen that our pairwise registration algorithm is very accurate. Most of the registered range image pairs have a rotation error less than 1.0° and a translation error less than $1.0d_{res}$.



(a) Rotation error (b) Translation error

Fig. 4: Histograms of the rotation and transformation errors of the four objects.

3) Robustness to the Degree of Overlap: We tested the performance of our algorithm with respect to varying degrees of overlap between range images to be registered. The degree of overlap between any two range images is known a priori. The numbers of correct and incorrect registrations of the four individual objects are shown in Fig. 5. Our algorithm achieved consistent results on the four individual objects. Generally, all the range image pairs with an overlap more than 60% were correctly registered (with rotation errors less than 5° and translation errors less than $5d_{res}$). With an overlap between 30% and 60%, about 75% of range image pairs were correctly registered. Moreover, correct registrations can even be achieved by several range image pairs with an overlap less than 20%, as shown in Fig. 5(b-d).



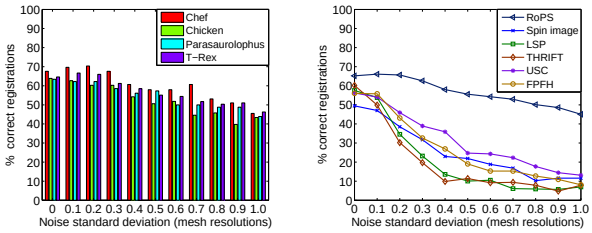
(a) Chef (b) Chicken

(c) Parasaurolophus (d) T-Rex

Fig. 5: Robustness to the degree of overlap.

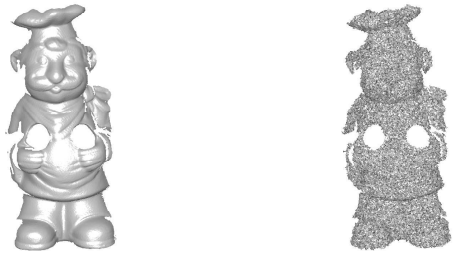
4) Robustness to Noise: We added a Gaussian noise to each mesh along the x, y and z axes. The standard deviation of noise was increased from 0 to $1.0d_{res}$ (d_{res} is the average mesh resolution) with an incremental step of $0.1d_{res}$. We then used our RoPS based algorithm to conduct pairwise registration on all range image pairs. The percentage of correct registrations of the four individual objects with respect to different levels of noise are presented in Fig. 6(a). Our algorithm is clearly very robust to noise, the percentage of correct registrations was almost unaffected by noise for meshes with a noise standard deviation less than $0.5d_{res}$. The algorithm still achieved acceptable results even for meshes with a noise deviation of $1.0d_{res}$. Note that, the surface of a mesh with a noise standard deviation of $1.0d_{res}$ is very spiky, and most of its shape details are lost (as shown in Fig. 7). We also compared our RoPS based algorithm to the spin image, LSP, THRIFT, USC and FPFH based algorithms. The combined results of all four objects achieved by different algorithms are shown in Fig. 6(b). It can be seen that our RoPS based algorithm achieved the best results at all levels of noise. The USC based algorithm achieved the second best overall performance. In contrast, both LSP and THRIFT based algorithms were very sensitive to noise. Their performance deteriorated rapidly as the standard deviation of noise increased. That is

because both LSP and THRIFT features rely on the surface normals or shape index values, which are very susceptible to noise as they require a process of surface differentiation.



(a) RoPS based algorithm (b) Comparison

Fig. 6: Pairwise registration performance with respect to different levels of noise (Figure best seen in color).

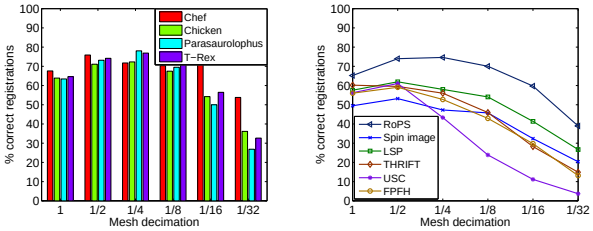


(a) (b)

Fig. 7: An illustration of a range image of the Chef. (a) A noise-free mesh. (b) A mesh with noise standard deviation of $1.0d_{res}$.

5) Robustness to Varying Mesh Resolutions: We simplified each mesh to varying mesh resolutions such that the number of points in a simplified mesh was $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$ and $\frac{1}{32}$ of its original value. The average vertex counts of Chef, Chicken, Parasaurolophus, and T-Rex meshes are 67737, 19019, 26193, and 31765, respectively. We tested our RoPS based algorithm with respect to varying mesh resolutions, the results for the four individual objects are shown in Fig. 8(a). Our RoPS based algorithm is shown to be very robust to varying mesh resolutions. The percentage of correct registrations with $\frac{1}{16}$ of original mesh resolution was even comparable with the results achieved on the original meshes. The performance started to drop when the simplified meshes had less than $\frac{1}{32}$ of their original number of vertices. We also compared our RoPS based algorithm to the spin image, LSP, THRIFT, USC and FPFH based algorithms. The combined results of all four objects achieved by different algorithms are shown in Fig. 8(b). The RoPS based algorithm outperformed the other algorithms at all levels of mesh resolution, followed by LSP, THRIFT and spin image based algorithms. The USC based algorithm was very sensitive to varying mesh resolutions. Its performance declined sharply when the simplified meshes had less than $\frac{1}{2}$ of their original number of vertices.

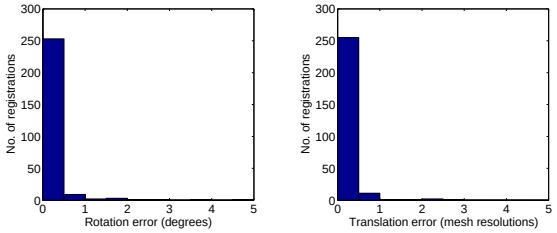
6) Results on a Synthetic Dataset: In order to further test our algorithm on range images for which the ground truth



(a) RoPS based algorithm (b) Comparison

Fig. 8: Registration performance with respect to varying mesh resolutions (Figure best seen in color).

transformations are known, we synthetically generated a set of range images from already built 3D models of the UWA dataset [11], in a way similar to [16]. Specifically, 24 range images were generated for each model from different viewpoints that are 15° apart in azimuth. We performed pairwise registration between any two range images of each object which have an overlap of more than 10%. The histograms of the rotation and transformation errors of the four objects are shown in Fig 9. It is clear that most of the registered range image pairs have a rotation error of less than 0.5° and a translation error of less than $0.5d_{res}$. The results on the synthetic dataset are more accurate compared to the results on the real dataset, as shown in Figs. 4 and 9. This is mainly due to the presence of noise (e.g., spikes) in the real range images.



(a) Rotation error (b) Translation error

Fig. 9: Histograms of the rotation and translation errors on synthetic range images of the four objects.

B. Multi-view Registration Results

In this section, we present the experimental results of our multi-view range image registration algorithm on the UWA 3D Modeling dataset.

1) Multi-view Registration of a Single Object: We used the range images of each object of the UWA 3D Modeling Dataset as an individual input, and tested the performance of our multi-view range image registration algorithm. Fig. 10 shows the range images and the multi-view coarse registration results of the Chicken and Parasaurolophus. Different range images are rendered in different colors. It can be observed that, although these range images were scanned from different viewpoints and presented in a random order, they were accurately registered. No visually noticeable defects or seams can be found in the registered range images (Fig. 10(b) and

(d)), even in the featureless parts of the objects (e.g., the tail of the Parasaurolophus in Fig. 10(d)).

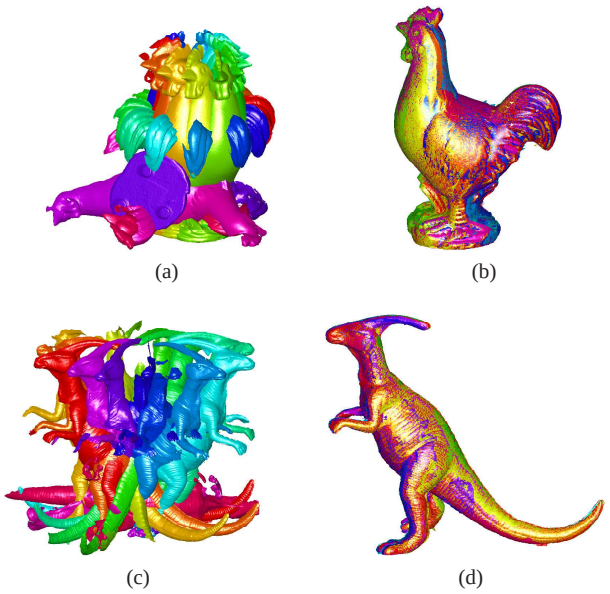


Fig. 10: An illustration of multi-view coarse registration results. (a) Range images of the Chicken. (b) Multi-view registration result of the Chicken. (c) Range images of the Parasaurolophus. (d) Multi-view registration result of the Parasaurolophus (Figure best seen in color).

In order to quantitatively analyze the accuracy of our multi-view coarse registration algorithm, we present the percentage of registered range images, and the average registration errors of each individual object in Table III(a). All range images of the four individual objects were correctly registered. The average rotation and translation errors of the four objects were less than 1.5° and $2.5d_{res}$, respectively. These yet accurate results were further refined by the subsequent fine registration algorithm; the results are shown in Table III(b). The average rotation and translation errors of the four objects were less than 0.5° and $0.5d_{res}$, respectively. Generally, our algorithm enables multi-view registration to be performed automatically and accurately.

TABLE III: Multi-view registration results of range images of four individual objects.

(a) Coarse registration

	Chef	Chicken	Parasaurolophus	T-Rex
#range images	22	16	16	21
%registration	100	100	100	100
Error ϵ_r ($^\circ$)	0.8559	1.1158	0.3456	0.5973
Error ϵ_t (d_{res})	0.9844	1.0595	1.3907	2.0051

(b) Fine registration

	Chef	Chicken	Parasaurolophus	T-Rex
#range images	22	16	16	21
%registration	100	100	100	100
Error ϵ_r ($^\circ$)	0.2712	0.3900	0.1771	0.3758
Error ϵ_t (d_{res})	0.3773	0.4508	0.0912	0.4161

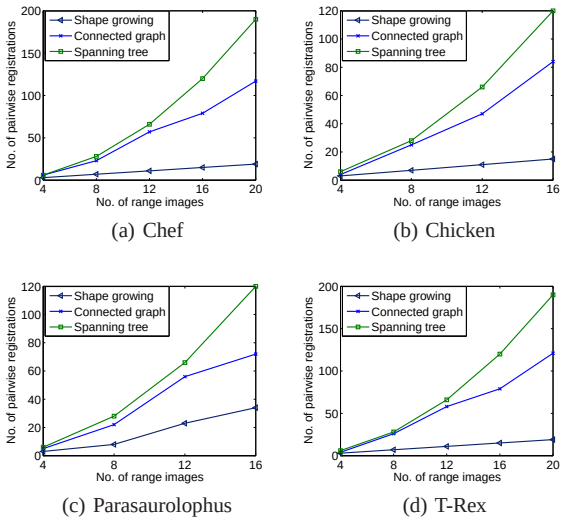
2) Robustness to Image Orders: In order to test the robustness of our multi-view registration algorithm with respect to different orders of the input range images, we randomly changed the order of input range images. We tested our algorithm on range images with five different orders. The fine registration results of the four objects are shown in Table IV. The results achieved with different image orders are very close to each other. The average rotation and translation errors were less than 0.6° and $0.6d_{res}$, respectively. We also present the results of the connected graph based algorithm [11] and the spanning-tree based algorithm [9] in Table IV. It is clear that our multi-view registration algorithm outperformed the connected graph based algorithm.

TABLE IV: Multi-view fine registration results on range images.

	%registration	Error ϵ_r ($^\circ$)	Error ϵ_t (d_{res})
Proposed (order 1)	100	0.4649	0.4854
Proposed (order 2)	100	0.3888	0.3922
Proposed (order 3)	100	0.3035	0.3339
Proposed (order 4)	100	0.5346	0.5182
Proposed (order 5)	100	0.4824	0.3909
Connected graph	100	0.5457	0.7477
Spanning tree	100	0.2796	0.2536

3) Efficiency w.r.t. the Number of Input Meshes: In order to evaluate the computational efficiency of the multi-view registration algorithm with respect to the number of input meshes, we progressively selected a subset of the range images to perform multi-view registration. For each fixed number of input meshes, we counted the number of pairwise registrations which were needed to complete the multi-view registration. The results for each of the four objects are shown in Fig. 11. We also present the results of the state-of-the-art including the spanning-tree based algorithms [9], [17], and the connected graph based algorithm [11]. Our shape growing based algorithm showed a significant improvement compared to both the spanning tree based and the connected graph based algorithms. Taking the 20 input range images of the Chicken as an example, the numbers of pairwise registrations for the spanning tree based, connected graph based and shape growing based algorithms were 190, 117 and 19, respectively. The improvement factor of our shape growing based algorithm over the spanning tree based algorithm was $\frac{190}{19} = 10$. Note that, as the number of input range images increases, the advantage of our algorithm becomes even more significant. We also measured the processing time to register all range images of each object. The timing experiments were conducted on a computer with a 3.5 GHz Intel Core i7 CPU and a 16GB RAM. The code was implemented in MATLAB. The average computational times were 21.82min, 7.39min, 24.54min and 17.17min for Chef, Chicken, Parasaurolophus and T-Rex, respectively. It should be noted that the running speed can further be improved and optimized by implementing the algorithms in C++.

4) Multi-view Registration of Multiple Objects: In order to further demonstrate the capability of our algorithm to simultaneously register multiple mixed range images corresponding to multiple objects, we used all the range images of the four objects at the input. These range images were



(a) Chef (b) Chicken

(c) Parasaurolophus (d) T-Rex

Fig. 11: Efficiency with different number of input meshes (Figure best seen in color).

mixed and were registered using our shape growing based algorithm. As a result, four shapes were produced by our algorithm. The totality of the 75 input range images are shown in Fig. 12(a), and the coarse registration results for the four shapes are respectively shown in Fig. 12(b-e). It can be seen that, all these input range images were separately registered according to their corresponding shapes. Moreover, although fine registration was not applied to these registration results, there were no visually noticeable seams in any of the registered range images.

We also present the percentage of correctly registered range images, and the average registration errors of each individual object in Table V. These results were almost the same as those reported in Table III. This observation clearly indicates that the mixture of range images from multiple objects has very few effect on the registration result of each object. This is due to the reason that, for a specific shape, the range images from different objects cannot be successfully registered with it. Consequently, these range images have no contribution to the final registration results. Generally, our algorithm is able to perform multi-view registration correctly from a mixed and unordered set of range images from several different objects.

C. 3D Object Modeling

We tested the 3D object modeling framework on both high-resolution and low-resolution datasets.

1) Results on High-Resolution Range Images: We first tested our 3D modeling framework on high-resolution range images from two popular datasets: the UWA 3D Modeling Dataset [11] and the Stanford 3D Scanning Repository [46]. The former consists of a set of range images of the Chef, Chicken, Parasaurolophus and T-Rex. The latter consists of a set of range images of the Armadillo, Bunny, Dragon and Happy Buddha. The number of range images of the Armadillo, Dragon and Happy Buddha are more than 50. The multi-view registration results and reconstructed 3D models of these

TABLE V: Multi-view registration results of mixed range images of the four objects.

(a) Coarse registration

	Chef	Chicken	Parasaurolophus	T-Rex
#range images	22	16	16	21
%registration	100	100	100	100
Error ϵ_r ($^\circ$)	1.8330	1.6183	1.2157	1.4149
Error ϵ_t (d_{res})	1.2674	1.3967	1.6750	1.8516

(b) Fine registration

	Chef	Chicken	Parasaurolophus	T-Rex
#range images	22	16	16	21
%registration	100	100	100	100
Error ϵ_r ($^\circ$)	0.3341	0.4286	0.4131	0.2369
Error ϵ_t (d_{res})	0.3975	0.3969	0.5258	0.1333

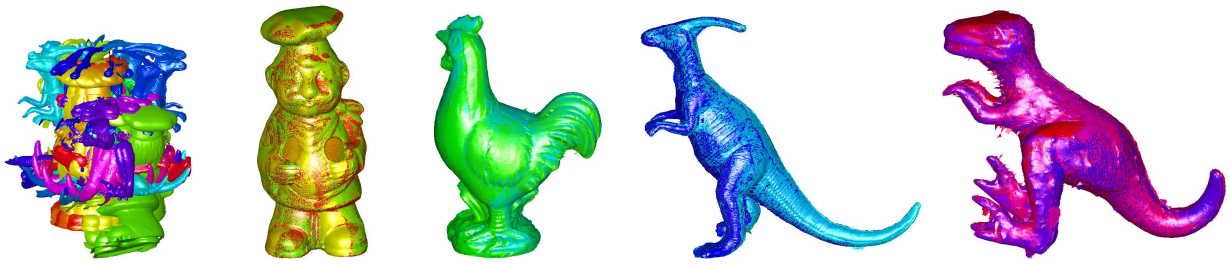
objects are shown in Fig. 13. These results clearly demonstrate that our algorithm is capable of reconstructing 3D models by seamlessly merging multiple range images.

In order to quantitatively analyze the performance of our 3D modeling framework, we compared our reconstructed models with the ground truth models. We use the term *accuracy* and *completeness* (proposed by Seitz et al. [48]) to evaluate our modeling results. To measure the *accuracy*, we calculated the distance d_{ac} such that 90% of the points on the reconstructed model are within the distance d_{ac} to the ground truth model. The distance d_{ac} was further normalized by the mesh resolution d_{res}. To measure the *completeness*, we calculated the percentage of points on the ground truth model that are within 2 times the mesh resolution to the reconstructed model. The accuracy and completeness results of the 8 reconstructed models are shown in Table VI. It is clear that our reconstructed 3D models are very accurate and complete compared to their corresponding ground truth models.

2) Results on Low-Resolution Range Images: We further tested our 3D modeling framework on low-resolution range images from the Bologna Reconstruction Dataset [18]. The range images of Duck and Frog were acquired with a Microsoft Kinect sensor, while those of Squirell and Mario were acquired with a Microsoft Space Time sensor. The low resolution and high noise level of these range images make the task of 3D modeling even more challenging. The input range images and the reconstructed models are shown in Fig. 14. The upper-left plot of each block corresponds to the input range images, with the other three plots corresponding to the reconstructed model observed from three different views. It is clear that our framework was able to reconstruct the 3D shape of an object from noisy and low-resolution range images without any manual intervention or assumption about the initial poses.

VI. Conclusion

In this paper, we have presented an accurate and robust algorithm for both pairwise and multi-view range image registration. A pairwise range image registration algorithm is proposed by integrating four modules including RoPS feature extraction, feature matching, robust transformation estimation and fine registration. We performed extensive experiments to assess the



(a) Input range images (b) Chef (c) Chicken (d) Parasaurolophus (e) T-Rex

Fig. 12: Multi-view coarse registration of range images corresponding to multiple objects (Figure best seen in color).

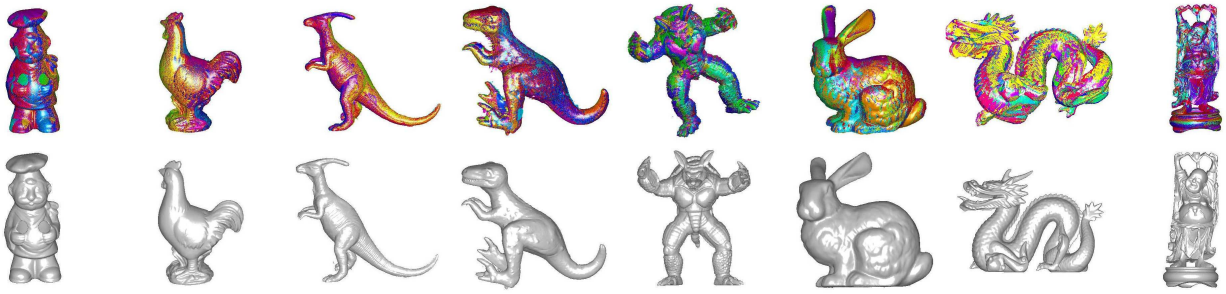


Fig. 13: 3D modeling results on high-resolution datasets (Figure best seen in color).

TABLE VI: Accuracy and completeness results of the reconstructed models.

	Chef	Chicken	Parasaurolophus	T-Rex	Armadillo	Bunny	Dragon	Happy Buddha
Accuracy (d_{res})	0.6231	1.1004	0.6942	0.9485	1.7977	1.4291	1.1545	0.7209
Completeness (%)	100.00	100.00	99.66	99.97	91.12	99.88	98.46	100.00

Fig. 14: 3D modeling results on low-resolution datasets (Figure best seen in color).

accuracy and robustness of our algorithm with respect to a set of nuisances including small overlaps, noise, and varying mesh resolutions. Comparative experimental results show that our RoPS based algorithm outperforms the state-of-the-art. We also propose a shape growing based algorithm for multi-view range image registration. Experimental results show that the proposed algorithm is very accurate. It can simultaneously perform multi-view registration on a set of mixed range images which correspond to several different objects. Finally, we introduced a complete 3D modeling framework based on our registration algorithms. Experimental results on both high-resolution and low-resolution range images show that the reconstructed 3D models are complete and accurate. In order to improve the processing time of our algorithm for real-life applications, our future work will aim to implement the proposed algorithms in a Graphics Processing Unit (GPU) with parallel computing techniques to achieve a faster performance.

REFERENCES

[1] Y. Gao, M. Wang, Z.-J. Zha, Q. Tian, Q. Dai, and N. Zhang, "Less is more: efficient 3-D object retrieval with query view selection," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 1007–1018, 2011.

[2] D. Alexiadis, D. Zarpalas, and P. Daras, "Real-time, full 3-D reconstruction of moving foreground objects from multiple consumer depth cameras," *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 339–358, 2013.

[3] T. Weise, T. Wismer, B. Leibe, and L. Van Gool, "In-hand scanning with online loop closure," in *IEEE 12th International Conference on Computer Vision Workshops*, 2009, pp. 1630–1637.

[4] G. Stavropoulos, P. Moschonas, K. Moustakas, D. Tzovaras, and M. G. Strintzis, "3-D model search and retrieval from range images using salient features," *IEEE Transactions on Multimedia*, vol. 12, no. 7, pp. 692–704, 2010.

[5] Y. Guo, J. Wan, M. Lu, and W. Niu, "A parts-based method for articulated target recognition in laser radar data," *Optik*, vol. 124, no. 17, pp. 2727–2733, 2013.

[6] T. Weise, B. Leibe, and L. Van Gool, "Accurate and robust registration for in-hand modeling," in *IEEE Conferenceon Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[7] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison *et al.*, "Kinectfusion: real-time 3D reconstruction and interaction using a moving depth camera," in *24th annual ACM symposium on User interface software and technology*, 2011, pp. 559–568.

[8] D. Thomas and A. Sugimoto, "Robustly registering range images using local distribution of albedo," *Computer Vision and Image Understanding*, vol. 115, no. 5, pp. 649–667, 2011.

[9] D. Huber and M. Hebert, "Fully automatic registration of multiple 3D data sets," *Image and Vision Computing*, vol. 21, no. 7, pp. 637–650, 2003.

[10] J. Salvi, C. Matabosch, D. Fofi, and J. Forest, "A review of recent range image registration methods with accuracy evaluation," *Image and Vision Computing*, vol. 25, no. 5, pp. 578–596, 2007.

[11] A. Mian, M. Bennamoun, and R. Owens, "Three-dimensional model-based object recognition and segmentation in cluttered scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1584–1601, 2006.

[12] C. Dorai, G. Wang, A. K. Jain, and C. Mercer, "Registration and integration of multiple object views for 3D model construction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 83–89, 1998.

[13] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 433–449, 1999.

[14] 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1105–1120, 2002.

[15] P. Bariya, J. Novatnack, G. Schwartz, and K. Nishino, "3D geometric scale variability in range images: Features and descriptors," *International Journal of Computer Vision*, vol. 99, no. 2, pp. 232–255, 2012.

[16] A. Mian, M. Bennamoun, and R. A. Owens, "A novel representation and feature matching algorithm for automatic pairwise registration of range images," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 19–40, 2006.

[17] T. Masuda, "Log-polar height maps for multiple range image registration," *Computer Vision and Image Understanding*, vol. 113, no. 11, pp. 1158–1169, 2009.

[18] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *European Conference on Computer Vision*, 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," *International Journal of Computer Vision*, vol. 105, no. 1, pp. 63–86, 2013.

[20] G. Dalley and P. Flynn, "Pair-wise range image registration: a study in outlier classification," *Computer Vision and Image Understanding*, vol. 87, no. 1, pp. 104–115, 2002.

[21] F. Stein and G. Medioni, "Structural indexing: Efficient 3D object recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 125–145, 1992.

[22] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *8th European Conference on Computer Vision*, 2004, pp. 224–237.

[23] F. Tombari, S. Salti, and L. Di Stefano, "Unique shape context for 3D data description," in *ACM Workshop on 3D Object Retrieval*, 2010, pp. 57–62.

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

[25] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, "Aligning point cloud views using persistent feature histograms," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008, pp. 3384–3391.

[26] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *IEEE International Conference on Robotics and Automation*, 2009, pp. 3212–3217.

[27] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan, "3D object recognition in cluttered scenes with local surface features: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence. Under minor revision*, 2013.

[28] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 5, pp. 698–700, 1987.

[29] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," in *IEEE International Conference on Robotics and Automation*, 1991, pp. 2724–2729.

[30] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[31] B. He, Z. Lin, and Y. Li, "An automatic registration algorithm for the scattered point clouds based on the curvature feature," *Optics & Laser Technology*, 2012.

[32] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *3rd International Conference on 3-D Digital Imaging and Modeling*, 2001, pp. 145–152.

[33] Y. Liu, "Automatic range image registration in the markov chain," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 12–29, 2010.

[34] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan, "TriSI: A distinctive local surface descriptor for 3D modeling and object recognition," in *8th International Conference on Computer Graphics Theory and Applications*, 2013, pp. 86–93.

[35] F. B. ter Haar and R. C. Veltkamp, "Automatic multiview quadruple alignment of unordered range scans," in *IEEE International Conference on Shape Modeling and Applications*, 2007, pp. 137–146.

[36] R. Benjemaa and F. Schmitt, "A solution for the registration of multiple 3D point sets using unit quaternions," in *European Conference on Computer Vision*. Springer, 1998, pp. 34–50.

[37] P. J. Neugebauer, "Reconstruction of real-world objects via simultaneous registration and robust combination of multiple range images," *International Journal of Shape Modeling*, vol. 3, no. 01n02, pp. 71–90, 1997.

[38] J. Williams and M. Bennamoun, "Simultaneous registration of multiple corresponding point sets," *Computer Vision and Image Understanding*, vol. 81, no. 1, pp. 117–142, 2001.

[39] T. Masuda, "Registration and integration of multiple range images by matching signed distance fields for object shape modeling," *Computer Vision and Image Understanding*, vol. 87, no. 1, pp. 51–65, 2002.

[40] K. Nishino and K. Ikeuchi, "Robust simultaneous registration of multiple range images," in *Asian Conference on Computer Vision*, 2002, pp. 454–461.

[41] G. Guennebaud and M. Gross, "Algebraic point set surfaces," *ACM Transactions on Graphics*, vol. 26, no. 3, p. 23, 2007.

[42] Y. Guo, M. Bennamoun, F. Sohel, J. Wan, and M. Lu, "3D free form object recognition using rotational projection statistics," in *IEEE 14th Workshop on the Applications of Computer Vision*, 2013, pp. 1–8.

[43] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[44] H. Du, P. Henry, X. Ren, M. Cheng, D. B. Goldman, S. M. Seitz, and D. Fox, "Interactive 3D modeling of indoor environments with a consumer depth camera," in *13th international conference on Ubiquitous computing*, 2011, pp. 75–84.

[45] D. A. Forsyth and J. Ponce, "Computer vision: A modern approach," *Prentice Hall*, 2002.

[46] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *23rd Annual Conference on Computer Graphics and Interactive Techniques*, 1996, pp. 303–312.

[47] S. Malassiotis and M. Strintzis, "Snapshots: A novel local surface descriptor and matching algorithm for robust 3D surface alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1285–1290, 2007.

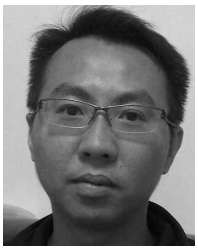
[48] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *IEEE Conference on Computer vision and pattern recognition*, vol. 1, 2006, pp. 519–528.

Yulan Guo received his B.Eng. in communication engineering from National University of Defense Technology (NUDT) in 2008. He is currently working towards his Ph.D. degree at NUDT. He has been a visiting (joint) Ph.D. student at the University of Western Australia from November 2011 to November 2013. He is a reviewer for several international journals including IEEE Transactions on Multimedia, IEEE Signal Processing Letters, and Neurocomputing. His research interests include 3D object recognition, 3D face recognition, 3D modeling, pattern recognition, and signal processing.

Ferdous Sohel received BSc degree in Computer Science and Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh (in 2002) and PhD degree from Monash University, Australia (in 2008). He is currently a Research Assistant Professor at the University of Western Australia. He has served on the program committee of a number of international conferences including an ICCV workshop. He is a recipient of prestigious Discovery Early Career Research Award (DECRA) funded by the Australian Research Council. He is a recipient of Mollie Holman Doctoral Medal and Faculty of Information Technology Doctoral Medal form Monash University. His research interests include computer vision, image processing, pattern recognition, and shape coding. He has published over 40 scientific articles.

Mohammed Bennamoun received his MSc degree in control theory from Queens University, Kingston, Canada, and his Ph.D. degree in computer vision from Queens/QUT in Brisbane, Australia. He is currently a Winthrop Professor at the University of Western Australia, Australia. He served as a guest editor for a couple of special issues in International journals such as the International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI). He was selected to give conference tutorials at the European Conference on Computer Vision (ECCV) and the International Conference on Acoustics Speech and Signal Processing (ICASSP). He organized several special sessions for conferences, e.g., the IEEE International Conference in Image Processing (ICIP). He also contributed in the organization of many local and international conferences. His research interests include control theory, robotics, obstacle avoidance, object recognition, artificial neural networks, signal/image processing, and computer vision. He published more than 200 journal and conference publications.

Jianwei Wan received his PhD degree from National University of Defense Technology (NUDT) in 1999. He is currently a professor with the College of Electronic Science and Engineering at NUDT. He has been a senior visiting scholar at the University of Windsor, Canada from May 2008 to May 2009. He published over 100 journal and conference publications. His research interests include signal processing, remote sensing, computer vision, pattern recognition, and hyperspectral image processing.



Min Lu received his B.Eng. and PhD degrees from National University of Defense Technology (NUDT) in 1997 and 2003, respectively. He is currently an associate professor with the College of Electronic Science and Engineering at NUDT. He is also a visiting scholar at the University of Waterloo, Canada. His research interests include computer vision, pattern recognition, remote sensing, signal and image processing.