Neurocomputing 000 (2017) 1–14

Contents lists available at ScienceDirect

Neurocomputing

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

Multi-attribute statistics histograms for accurate and robust pairwise registration of range images

Jiaqi Yang, Qian Zhang, Zhiguo Cao*

Guangdong HUST Industrial Technology Research Institute, Guangdong Province Key Lab of Digital Manufacturing Equipment, Image Processing and Intelligent Control Key Laboratory of Education Ministry of China, School of Automation, Huazhong University of Science and Technology, Wuhan, China

ARTICLE INFO

Article history:
Received 8 August 2016
Revised 20 January 2017
Accepted 5 April 2017
Available online xxx

Communicated by Bin Fan

Keywords:
Local reference axis
Multi-attribute
Transformation estimation
Feature matching
Range image registration

ABSTRACT

Registration of range images based on local shape features is widely adopted due to its validated effectiveness and robustness, while most existing local shape descriptors struggle to simultaneously achieve a pleasurable and balanced performance in terms of distinctiveness, robustness and time efficiency. This paper proposes a novel representation of 3D local surfaces, called multi-attribute statistics histograms (MaSH), for automatic registration of range images. MaSH comprises both spatial and geometric information characterizations. The characterization of spatial information is achieved via radial partitions in the 3D local support volume around the keypoint based on a local reference axis (LRA), creating a set of subspaces. While the encoding the shape geometry is performed by calculating statistical histograms of multiple faint correlated geometric attributes (i.e., local depth, normal deviation, and surface variation angle) for each subspace, so as to achieve information complementarity. Then, a robust rigid transformation estimation algorithm named 2-point based sample consensus with global constrain (2SAC-GC) is presented to tackle the problem of calculating an optimal transformation from the correspondence set with outliers. Finally, a coarse-to-fine registration method based on MaSH and 2SAC-GC is proposed for aligning range images. Experiments on both high-resolution (Laser Scanner) and low-resolution (Kinect) datasets report that, our method achieves a registration accuracy of 90.36% and 80.39% on the two datasets, respectively. It also exhibits strong robustness against noise and varying mesh resolutions. Furthermore, feature matching experiments show the over-all superiority of the proposed MaSH descriptor against the state-of-the-arts including the spin image, snapshots, THRIFT, FPFH, RoPS, LFSH and RCS descriptors.

1. Introduction

There has been a long research history on the computer vision tasks based on 2D images. Despite their notable progre ss, there are still numerous obstacles left due to the essential shortcomings of 2D images, e.g., sensitivities to scale, rotation and illumination. In contrast, range images, as the 3D representation of the real world, show many favorable traits against 2D images, such as providing more geometrical (depth) information and scale invariance. Accordingly, vision tasks based on range images have attracted growing interests. The development of numerous low-cost 3D sensors, e.g., Microsoft Kinect and Intel RealSense, has further promoted such trend in recent years. Pairwise registration of range images is a fundamental and crucial step in many 3D vision tasks, such as 3D object recognition [1], surface registration [2], 3D modeling [3], cultural heritage [4] and robotics [5].

Given two range images, namely a source one and a target one, the aim of pairwise range image registration is to transform the coordinate of the source to that of the target. The challenges are many fold, e.g., unknown initial poses, noise injected by the acquisition system, varying mesh resolutions (point densities) caused by different scanning distances, holes and small overlaps. Some applications, such as robotics, even have strict demands on time efficiency. To solve above difficulties, many efforts have been made in the literature. A matured pipeline for range image registration is a coarse-to-fine strategy [3,6–8], where the coarse registration provides an initial guess for the fine registration as well as accelerates the registration process. The most typical fine registration methods are arguably the iterative closest point (ICP) algorithm and its variants [9–11]. Since the existing fine registration methods are already able to obtain fairly accurate results [12], the real challenging turns out to be the coarse registration problem, which has attracted many interests lately. Feature-based

* Corresponding author.
E-mail addresses: jqyang@hust.edu.cn (J. Yang), hangfanzq@163.com (Q. Zhang), zgcao@hust.edu.cn (Z. Cao).

http://dx.doi.org/10.1016/j.neucom.2017.04.015
0925-2312/

coarse registration methods are mostly used nowadays owing to their validated effectiveness and robustness [3,6,13–17], their employed features can be classified into local and global features. Owing to the superiority of local features against global features in terms of distinctiveness and robustness to clutter, occlusion, and missing regions [12], local features are more suitable for generating correspondences between range images.

For local feature-based methods, local 3D descriptor and transformation estimation play two critical roles because the former one directly affects the quality of the established correspondences between two range images, and the later one is amenable to calculate a reasonable transformation from a correspondence set with outliers. In the following, we briefly review the existing local feature descriptors and 3D transformation estimation methods in the literature. Johnson and Hebert [6,18] proposed a spin image descriptor for surface matching and object recognition. They represented the neighboring points of a keypoint with a cylindrical coordinate frame, and encoded the point distribution with a 2D image, named a spin image. Spin image is a one of the most cited descriptor, but still suffers from limited descriptiveness and poor robustness. Several variants of spin image are later proposed, including the spherical spin image [19] and Tri–Spin-Image (TriSI) descriptor [20]. Frome et al. [21] proposed a 3D shape context (3DSC) descriptor, which is an extension of the 2D shape context method [22]. They first divided the local spherical volume around the keypoint into a set of subspaces based on a local reference frame (LRF), and then calculated the weighted number of points in each subspace. Flint et al. [15,23] first calculated the normal deviations between the normals of the neighbors and that of the keypoint, and then represented the normal deviation distribution with a 1D statistical histogram, called THRIFT. Similarly, Rusu et al. [7,8] proposed the point feature histograms (PFH) and fast point feature histograms (FPFH) for characterizing normal attributes. Malassiotis and Strintzis [24] first built an LRF centered at the keypoint by performing covariance analysis on the neighboring points, they then captured a local depth image at a virtual view point in the z-axis of the LRF, generating a “snapshots” descriptor. Guo et al. [25] proposed a rotational projection statistics (RoPS) descriptor, which is the concatenation of the statistics calculated for each 2D projection map after rotating the local surface multiple times. Following the multi-view mechanism in RoPS, a rotation contour signature (RCS) [26] based on 2D contour feature representation is proposed in [26], delivering better robustness than RoPS. Shah et al. [27,28] proposed a 3D vorticity (3D-Vor) descriptor based on the vorticity of the normal vector field computed at each neighboring point. 3D-Vor particularly addresses low-resolution range image registration scenario. Recently, Yang et al. [29] proposed a local feature statistics histograms (LFSH) descriptor to achieve fast point cloud registration using several low-dimensional and efficient geometrical features. Although there has been a variety of descriptors at present, most of them still fail to achieve a balanced and pleasant performance. For instance, spin image and LFSH exhibit limited descriptiveness [30]; PFH and FPFH are sensitive to noise [3]; snapshots is prone to rotation [25]; RoPS demands mesh representation and is computational expensive [31]. However, these are all major concerns for a local feature-based registration algorithm in practice.

In terms of transformation estimation, the random sample consensus (RANSAC) [32] is a commonly-used method in both 2D and 3D registration cases, and shows some specific advantages, e.g., low implementation complexity and very few tunable parameters, against other approaches such as 3D Hough transform [33], the rigidity constrain framework [34] and voting-based approach [35]. In 3D registration case, owing to that additional geometry information and/or constrains can be explored, many variants of RANSAC are thereupon proposed to either improve its time efficiency and/or robustness. Rusu et al. [8] proposed a sample consensus initial alignment (SAC-IA) algorithm by selecting a set of sample correspondences with distance constrain from the initial correspondence set, and iteratively computing the optimal transformation that yields to best error metric (i.e., the Huber penalty measure). The SAC-IA algorithm reduces some unreasonable iterations by judging the qualification of current sampled correspondences according to a distance constrain, while it still struggles to take effect in cases with severe outliers. Guo et al. [36] recently proposed a 1-point RANSAC (1P-RANSAC) algorithm to reduce the high computational complexity of RANSAC. To be specific, they first calculated an LRF for each keypoint in both range images, and then iteratively sampled a single correspondence from the correspondence set for the calculation of the optimal transformation. Owing to that two corresponding LRFs are suffice to compute a rigid transformation [37], the 1P-RANSAC algorithm only needs one correspondence at each iteration in opposite to three in the classical RANSAC algorithm. However, it is sensitive to LRF calculation errors and high percentage of outliers in the initial correspondence set.

In these regards, we propose a novel local shape descriptor called multi-attribute statistics histograms (MaSH) together with a new transformation estimation algorithm named 2-point based sample consensus with global constrain (2SAC-GC), for accurate and robust pairwise registration of range images. MaSH contains both spatial and geometric information descriptions. The spatial information is encoded via partitions along the radial direction in the local 3D support volume based on an LRA. Employing 1D spatial information is a compromise among descriptiveness, robustness and efficiency, since more dimensional spatial information (e.g., the azimuth dimension) requires an LRF rather than an LRA. However, establishing a repeatable LRF in the local surface is far more challenging (i.e., needs to calculate three unique axes rather than one) and needs more time consumptions than an LRA [38,39]. As for geometric information description, statistics histogram of point attribute is a simple and effective feature representation as in [8,15,30]. Specifically, we calculate multiple (three in this paper) faint correlated point attributes for each neighboring point, and generate three statistic histograms for each subspace (partition). The idea behind multi-attribute description is to give a comprehensive geometric information encoding for the local surface, in opposite to traditional single attribute-based methods [15,18,25]. By concatenating these histograms in each subspace into a vector, a MaSH is generated. Then, we present a transformation estimation algorithm named 2-point based sample consensus with global constrain (2SAC-GC), which iteratively calculates the optimal transformation from two correspondences with two geometry constrains and a new discriminant criterion. Finally, a coarse-to-fine method is proposed for automatic pairwise registration of range images.

With the same purpose of local surface description as in our previous works [26,29], though, the designing techniques in the proposed MaSH descriptor are different and more advanced. First, compared with [29], MaSH integrates spatial information and histogram-weights to attain stronger discriminative power and resilience to self-occlusion. Second, in opposite to [26], MaSH is a histogram-based descriptor while [26] is signature-based. Further, MaSH does not rely on LRF for spatial information characterization. Third, the crafting scheme in MaSH particularly addresses the range image registration scenario, i.e., self-occlusion, missing region and holes, whereas [26,29] mainly consider complete local shapes. Comparative results (Section 4.3) clearly demonstrate that the proposed MaSH descriptor outperforms our previous works [26,29] on all the datasets in both feature matching and range image registration performance.