Pattern Recognition 66 (2017) 375–391

Contents lists available at ScienceDirect

Pattern Recognition

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

The effect of spatial information characterization on 3D local feature descriptors: A quantitative evaluation

Jiaqi Yang, Qian Zhang, Zhiguo Cao*

National Key Laboratory of Science and Technology on Multi-spectral Information Processing, School of Automation, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, PR China

ARTICLE INFO

Keywords:
Local feature descriptor
Spatial information
Local reference axis
Local reference frame
Feature matching

ABSTRACT

Designing local feature descriptors for 3D objects is a fundamental yet challenging task in 3D computer vision. Both geometry and spatial information descriptions are critical for a 3D local descriptor, while most previous studies concentrate on the former one. This paper investigates on how the characterization of spatial information would affect a 3D local descriptor in terms of descriptiveness, robustness, compactness and efficiency. The evaluation process is deployed as follows. First, based on the analysis of representative spatial information characterization methods of existing local shape descriptors, six typical characterization methods with different spatial dimensions and partition principles of spatial information are presented. Second, three geometric attributes, i.e., normal deviation, local depth and shape index, are respectively assigned to each point in the local surface for local geometry description, creating a total of 18 different feature descriptors. Then, a quantitative analysis of performance (i.e., descriptiveness, robustness, compactness and efficiency) for these descriptors is carried out on three benchmark datasets. Grounded on the experimental outcomes, the traits, merits and demerits of each spatial information encoding approach are eventually summarized. This study reveals that different spatial information encoding approaches would bring significant effect on a local shape descriptor with respect to its discriminative power, stability, compactness and efficiency.

1. Introduction

In the last few years, the flourishment of low-cost 3D sensors, such as Microsoft Kinect, Intel RealSense and Google Project Tango, makes it quite popular to access the 3D data (point clouds, meshes and depth images). Such convenience has greatly boosted the development in 3D computer vision. Similar to the role played by 2D feature descriptors such as SIFT [1] and HOG [2] in 2D vision tasks, 3D local feature descriptors have manifested their significance in numerous applications, e.g., 3D registration [3], 3D object recognition [4], human face recognition [5], shape retrieval [6], and 3D remote sensing [7].

A local surface descriptor is designed for performing effective shape description for a small subset of 3D data and should be invariant to rigid transformations. Further, concerning nuisances like noise, varying mesh resolutions (point densities), clutter and occlusion, it should also hold strong robustness. Besides, some typical applications including robots and mobile phones also have strict limits on the compactness and efficiency of a feature descriptor. Above rules give rise to tremendous challenges of designing an effective and balanced descriptor. Many attempts have been made over the last two decades. Some researchers focused on exploring distinctive point attributes for representing the local shape geometry. For instance, Hetzel et al. [8] analyzed the descriptiveness of three point attributes, i.e., pixel depth, surface normals and curvature, by performing object recognition experiments using their 1D histogram representations. Rusu et al. [9,10] chose the normal attribute (i.e., deviation angle between normals) to describe the local shape geometry, and generated two statistical histograms called the point feature histograms (PFH) and the fast point feature histogram (FPFH). Yamany and Farag [11] employed the curvature information of the local surface for shape description, generating a 2D histogram named surface signature. Chen and Bhanu [12] combined shape indices and normal deviations of the neighboring points for feature representation, and proposed a local surface patch (LSP) descriptor. These descriptors are mainly composed of statistical histograms of various point attributes. Owing to the lack of spatial information, the resultant limited descriptiveness is their major drawback [13], though they are usually low-dimensional and computational efficient.

As later highlighted in [13,14], encoding the shape geometry together with the local spatial information would significantly improve

* 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.patcog.2017.01.017
Received 21 January 2016; Received in revised form 7 December 2016; Accepted 10 January 2017
Available online 16 January 2017
0031-3203/

the descriptiveness of a feature descriptor. This category of feature descriptors characterizes the spatial information via partitions in the local 3D volume either based on a local reference axis (LRA) or a local reference frame (LRF). What make LRA and LRF different are their structures. An LRF is composed of three orthogonal axes while an LRA comprises a single orientated axis. Hence, LRF provides the entire local 3D spatial information including radial, azimuth and elevation dimensions, whereas LRA has one degree of freedom in the azimuth direction [15]. For LRA-based descriptors [16–18], many of them select the radial dimension for spatial information encoding. Examples include the log-polar height map (LPHM) [17], which first defines a local log-polar coordinate frame at the keypoint via radial divisions, and then calculates the local depth attributes with respect to an LRA for the local points, resulting a 2D statistical map that is eventually transformed into a 1D descriptor vector. The recent colored histograms of spatial concentric surflet-pairs (CoSPAIR) [19] descriptor also splits the local spherical volume into several sub-spaces along the radial direction. For LRF-based descriptors [20,21,13,14,22], although three dimensions of spatial information are available, some of them simply use 2D spatial information. For instance, Zaharescu et al. [23] first computed the gradient attributes for the points underlying the local surface, and then projected these gradient vectors onto three orthonormal planes of an LRF. For each plane, segmentation was performed and each subregion was associated with an eight-element histogram, which was also a sub-feature of the final generated MeshHOG descriptor [23]. Guo et al. [14] calculated several signatures after rotating and projecting the local surface in an LRF multiple times, and then concatenated them into a rotational projection statistics (RoPS) descriptor. Both MeshHOG and RoPS integrate 2D spatial information in an LRF, leaving one dimensional freedom via 3D-to-2D projection. While other LRF-based descriptors perform both radial, azimuth and elevation divisions on the local 3D volume to achieve more detailed characterization for spatial information, this is the case of intrinsic shape signature (ISS) [24], 3D shape context (3DSC) [15] and signature of histograms of orientations (SHOT) [13,25]. All of them first segment the local 3D volume into multiple subspaces using partition along the radial, azimuth and elevation directions, and then perform attribute description in each subspace.

Above LRA and LRF based descriptors encode spatial information with various manners, while their merits or demerits remain unclear for two reasons. First, the employed point attributes of each descriptor for geometry description are different. Second, the LRAs/LRFs of these descriptors may also differ from each other. Although there has been several studies on the evaluation of 3D local features, e.g., the comparative researches in [26,27], they only tested the performance of different feature descriptors and the factor of spatial information encoding is not targetedly evaluated, while both geometric attributes, spatial information encodings and feature representations would affect the performance of a feature descriptor [4]. This motivates us to explore the specific impacts brought by different spatial information characterization methods on a 3D feature descriptor. In particular, our scope covers the major concerns of a feature descriptor including descriptiveness, robustness, compactness and efficiency.

In this work, we quantitatively evaluate the effects of different encoding manners for spatial information on 3D local descriptors. Specifically, six different 3D spatial partition approaches that contain different spatial dimensions and/or division principles are taken into consideration. These spatial information characterization methods cover the majority of existing ones, and also include some new attempts. As for local geometry description, we respectively assign three popular point attributes, i.e., normal deviation, local depth and shape index, to each point in the local surface. Note that although this paper aims at exploring the effect of spatial information encoding on local shape descriptors, it is necessary to test the effects of different attribute descriptions on the consistency of the final results. However, both the evaluation of the performance of geometric attributes and the combination of different spatial divisions and geometric attributes are beyond the main focus of this paper. As a result, a total of 18 local shape descriptors are constructed for our evaluation in this paper. To quantitatively justify the performance of these descriptors, comprehensive experiments are conducted on three standard datasets. We also summarize the advantages, shortcomings and suitable applications of different approaches for spatial information characterization. The contributions of this paper are threefold.

- We investigate 6 spatial information characterization methods, and thereupon present 18 feature descriptors (three point attributes per method) for local shape description.
- We quantitatively evaluate and compare the performance (i.e., descriptiveness, robustness, compactness and efficiency) of these descriptors on three benchmark datasets.
- Instructive summarizations including the traits, merits and demerits of different approaches for characterizing spatial information are presented.

The reminder of this paper is organized as follows. Section 2 provides the details of the presented 18 local surface descriptors with different encoding techniques for spatial information. Section 3 analyzes the parameters of feature descriptors. Experimental results and analysis are provided in Section 4. Section 5 gives a summarization of the traits, merits and demerits of different spatial information encoding approaches. The conclusion and future work are drawn in Section 6.

2. Methodology description

In this section, we first describe three widely-used point attributes, i.e., normal deviation [18,10], local depth [28,29] and shape index [30,12] for local geometry encoding. Then, we illustrate the concept of LRA and LRF, and briefly introduce the employed LRA and LRF in this paper. On these bases, six approaches for the characterization of spatial information are put forward, and eighteen feature descriptors are thereupon presented.

2.1. Local geometry information encoding

There are many effective point attributes for representing local geometry information, including curvature [8,11], normal deviation [18,10], local depth [28,29], shape index [30,12] and etc. When spatial information is dropped, feature descriptors can be generated using the statistics of one or several of these point attributes, e.g., THRIFT [18] and local feature statistics histogram (LFSH) [29]. The common property of these attributes is being invariant to rigid transformations, although their descriptiveness and robustness may vary from each other. Because the focus of this paper lies in the encoding of spatial information, we have no strict demand for the selection of point attributes. Specifically, we choose three reference attributes, i.e., normal deviation (nd), local depth (ld), and shape index (si) for local geometric encoding, which are respectively calculated as follows.

Let $\mathcal{P}$ be an input 3D model, and p be a keypoint in $\mathcal{P}$. The radius neighbors of p are calculated as $Q = \{q_i: \|q_i - p\| \le r\}$, where r is the radius of the neighboring spherical volume. Note that Q constitutes a local surface that will be later encoded by a feature descriptor. The normal deviation attribute [29] of a neighboring point q_i is calculated as:

$$nd(q_i) = \arccos(n_{q_i}, n_p), \tag{1}$$

where n_{q_i} and n_p respectively represent the normals of q_i and p, and $nd(q_i) \in [0, \pi]$. The calculation of point normals follows the method in [31], which computes a 3×3 covariance matrix $\mathbf{M} = \frac{1}{k}\sum_{i=1}^{k}[q_i - \overline{q}]\cdot[q_i - \overline{q}]^T$ for the local surface Q. Here, k denotes the point quantity of Q