Computers & Graphics 46 (2015) 117–129

Contents lists available at ScienceDirect

Computers & Graphics

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

SMI 2014

Local deep feature learning framework for 3D shape

Shuhui Bu [a], Pengcheng Han [a], Zhenbao Liu [a,*], Junwei Han [a], Hongwei Lin [b]

[a] Northwestern Polytechnical University, China
[b] Zhejiang University, China

ARTICLE INFO

Article history:
Received 8 July 2014
Received in revised form
30 August 2014
Accepted 17 September 2014
Available online 2 October 2014

Keywords:
Bag-of-words
Deep learning
Shape correspondence
Shape symmetry detection
Shape recognition
Local deep feature

ABSTRACT

For 3D shape analysis, an effective and efficient feature is the key to popularize its applications in 3D domain. In this paper, we present a novel framework to learn and extract local deep feature (LDF), which encodes multiple low-level descriptors and provides high-discriminative representation of local region on 3D shape. The framework consists of four main steps. First, several basic descriptors are calculated and encapsulated to generate geometric bag-of-words in order to make full use of the various basic descriptors' properties. Then 3D mesh is down-sampled to hundreds of feature points for accelerating the model learning. Next, in order to preserve the local geometric information and establish the relationships among points in a local area, the geometric bag-of-words are encoded into local geodesic-aware bag-of-features (LGA-BoF). However, the resulting feature is redundant, which leads to low discriminative and efficiency. Therefore, in the final step, we use deep belief networks (DBNs) to learn a model, and use it to generate the LDF, which is high-discriminative and effective for 3D shape applications. 3D shape correspondence and symmetry detection experiments compared with related feature descriptors are carried out on several datasets and shape recognition is also conducted, validating the proposed local deep feature learning framework.

1. Introduction

The advancement of modeling, digitizing, and visualizing techniques for 3D models has led to an increasing amount of 3D models in the fields of multimedia, graphics, virtual reality, amusement, design, and manufacturing [1]. Nowadays, a large number of publicly available models such as Google 3D Warehouse have been quickly spread online. In addition, with the development of RGB-D devices, e.g., Microsoft Kinect, users can obtain 3D models in a convenient and efficient way, which further leads to the explosion of 3D data. This rapid growth causes high demand of 3D models techniques including shape retrieval, recognition, classification, and correspondence [2].

Reviewing the implementations of these techniques, we can find that feature-based methods play an important role and some details could be found in an early work [1] and the latest works [3,4]. 3D shape descriptors are used to characterize important global or local geometric characteristics, which are distinctively discriminative with other shapes or local regions. Some descriptors, such as area and volume, shape distributions [5], ratios derived from the object's convex-hull, or the Light Field Descriptor [6] have been proposed and achieved great performance in the task of matching, retrieval, and some other applications [7]. Although these techniques make some breakthrough, there still remain many hard problems badly in need of being solved. For example, the shape descriptors mentioned above are global and just use a single vector to represent an object. However, 3D models have rich information including surface, color, and texture, while a single vector cannot represent an object effectively. In addition, global descriptors are often not invariant to scaling, rotation, or translation, hence they have limited capability to discriminate shape variance. To cope with above-mentioned problems, local descriptors [4,7] which use a single vector to describe the local surface region around a number of sample points on an object, have been studied in recent decades. They have the merits of capturing important geometric changes on local regions of 3D surface, being invariant to scaling, rotation, and isometric transformation.

A good local descriptor is the one that is invariant to "unimportant" geometric changes, especially rotation, translation, scaling, or bending (such as changing the pose of an articulated character) [8]. Because only considering the feature of point itself, the descriptor can be influenced easily by geometric changes. When extracting the local feature of a 3D shape, we should take into account the neighbor area surrounding the feature point. In the last decade, some local descriptors, which will be introduced in the following section in detail, have been proposed and successfully used in many tasks. However, the performance of

* Corresponding author.
E-mail address: liuzhenbao@nwpu.edu.cn (Z. Liu).

http://dx.doi.org/10.1016/j.cag.2014.09.007
0097-8493/

some 3D shape local descriptors is still far from satisfactory. The main issue results from three aspects: First, some local descriptors are insufficient to describe complex 3D shape, i.e., only catching a piece of geometric characteristics. Second, 3D shape is composed of complex topological structure and visibly variational geometry, consequently for one type of descriptor only limited information can be extracted. Third, although some descriptors can collect enough information about one local region of 3D shape, they are redundant, which leads to the inefficient usage. Thus, in order to make the extracted feature boost the performance of shape analysis, it is vital to design an effective and efficient local descriptor which can provide discriminative information from raw data.

In this paper, we propose a novel framework to learn and extract local descriptor for 3D shape. The fundamental of the framework is to extract an intermediate representation preserving its surrounding information from low-level 3D descriptors. However, this intermediate representation is redundant which results in the low efficiency. Recently, the deep learning [9–11] has been applied successfully in speech recognition, image processing, and so on. First of all, it can provide a powerful solution to get the high-level feature that is discriminative and robust. So it is critical for pattern recognition. In addition, this method can achieve better generalization because of that the high-level feature is learned from the low-level features. Thus, we adopt deep belief networks (DBNs) to extract compact feature from the intermediate representation. Through the unsupervised learning, model parameters of DBNs are optimized, and the output of the DBNs for newly input data is regarded as the high-level feature which is called as local deep feature (LDF).

The advantages of this framework are as follows:

1. The framework is not only limited to SI-HKS or AGD, other local descriptors are also supported.
2. Multiple features can be fused to provide abundant description.
3. The learning procedure is fully unsupervised.
4. Unlike other machine learning methods which need to tune parameter manually for obtaining the best performance, there are no parameters to be tuned in the learning procedure. Some other parameters, which are used for generating intermediate representation, have little influence on the performance and it is easy to select proper parameters.

Several experiments are conducted in 3D shape correspondence, symmetry detection, and shape recognition tasks. Results and comparisons with related descriptors indicate that the proposed framework reaches promising performance.

2. Related work

Extrinsic descriptors: Some local descriptors are extracted based on location and orientation of 3D mesh, or a local coordinate system defined on a vertex. An earlier and representative work is spin images [12]. Recently, Darom et al. [13] extend the spin images to possess the capability of scale-invariant and interest point detection. Sipiran et al. [14] adopt 3D Harris detector to locate interesting points for 3D shape retrieval, which can be seen as an extension from 2D Harris detector measuring the variation in the gradient of a given function (e.g., the intensity function of a image). 3D SURF descriptor [15,16] is recently proposed for classifying and retrieving similar shapes. Although these features have been applied in many 3D shape processing applications, they belong to the extrinsic descriptors and usually cannot preserve the rich information on local region of 3D shape.

Intrinsic descriptors. To overcome the above limitations, several intrinsic descriptors have been proposed in recent decades, which do not need to specify the descriptor position relative to an arbitrarily defined coordinate system. Therefore, they achieve much better discriminative capability for 3D shape analysis.

Laplace Beltrami operator, which is a generalization of the Laplacian from flat space to manifold, is appealing for 3D shape retrieval because of sparse, symmetric, and intrinsic properties of its robustness to rigid transformation and deformation. Retrieval methods [17–20] extract main eigenvalues and eigenvectors of Laplace matrix generated on local regions to match different regions of 3D shapes. Laplace–Beltrami operator also provides an efficient way of computing a conformal map from a manifold mesh to a homeomorphous surface with constant Gaussian curvature. The histogram of conformal factors [21] serves as a robust pose-invariant signature of 3D shape, which is regarded as an attribute of a graph node to identify segmented parts in bipartite graph matching for 3D shape retrieval [22]. In a recent work [23], 3D shape is also partitioned into several connected iso-surfaces (annuluses) of conformal factors, and expressed with a graph where node substitutes each annulus.

Heat kernel signature [24], a recently proposed local descriptor, absorbs researchers' much attention. It provides rich local geometric information which makes the signature invariant to isometric deformation and has multi-scale characteristics, thereby achieving better performance in 3D shape retrieval and matching [25–28]. In order to overcome the influence of diffusion time change under different shape scales [25], Fourier transform is imposed on heat kernel signature at each given vertex to obtain scale invariant. Another work uses intrinsic shape context (ISC) [29] to characterize the local shape property. In the method, the shape context is processed in an intrinsic local polar coordinate system, therefore it is intrinsic and invariant to isometric deformation. Furthermore, Fourier transform is applied to the original shape content data to deal with orientation ambiguity.

Learning features. Feature learning based methods attract attention of many researchers in the last decade because of their capability of improving discriminability of low-level feature.

In the research of Shape Google [27,28], despite the introduction of spatial-sensitive bag-of-features (SS-BoF), the authors also present a similarity-sensitive hashing method to achieve the best discriminability and compact representation. A middle-level feature extraction scheme through learning hidden states from local basic descriptors is proposed by Castellani et al. [30,31]. In the method, local patches are modeled as a stochastic process through a set of circular geodesic pathways and learned via hidden Markov model. Bu et al. [32] propose shift-invariant ring feature (SI-RF) based on iso-geodesic rings and shift-invariant sparse coding for 3D shape analysis. It represents the local region of a feature point efficiently and has great performance on correspondence and retrieval tasks.

The Laplacian-based descriptors achieve state-of-the-art performance, however, they usually focus on different properties of shape and are suitable for specified task. In order to provide a generic feature descriptor for 3D shape, Litman et al. [33] propose a learning scheme for the construction of optimized spectral descriptors. In order to collect rich information from the raw data and select the most significant feature, Barra et al. [34] propose a method utilizing multiple kernel learning to find optimal linear combination of kernels in classification and retrieval.

Above-mentioned methods focus on feature itself, but ignore the structure consistency. Structural learning, which can produce high-level semantic labels from low-level features through a global optimization, has been successfully applied to segmentation or labeling. Kalogerakis et al. [35] introduce a data-driven approach to simultaneous segmentation and labeling of parts in 3D meshes. They adopt conditional random field model with defined terms assessing the consistency of faces with labels and