Neurocomputing 151 (2015) 583–592

Contents lists available at ScienceDirect

Neurocomputing

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

Locality-constrained sparse patch coding for 3D shape retrieval

Zhenbao Liu, Shuhui Bu*, Junwei Han

Northwestern Polytechnical University, Post Box 727, Youyi West Road 127, Xi'an 710072, China

ARTICLE INFO

Article history:
Received 16 November 2013
Received in revised form
22 March 2014
Accepted 1 June 2014
Available online 4 November 2014

Keywords:
3D shape retrieval
Patch
Sparse coding
Locality constraint

ABSTRACT

3D shape retrieval is a fundamental task in many domains such as multimedia, graphics, CAD, and amusement. In this paper, we propose a 3D object retrieval approach by effectively utilizing low-level patches of 3D shapes, which are similar as superpixels in images. These patches are first obtained by means of stably over-segmenting 3D shape, and then we adopt five representative geometric features including shape diameter function, average geodesic distance, and heat kernel signature, to characterize these low-level patches. A large number of patches collected from shapes in a dataset are encoded into patch words by virtue of locality-constrained sparse coding under the consideration of local smooth sparsity. Input query is compared with 3D models in the dataset through probability distribution of patch words. Experiments reveal that the proposed method achieves comparable retrieval performance to state-of-the-art methods.

1. Introduction

3D model as an important media contains rich 3D information preserving real object surface, color, and texture, which has been extensively applied in the domain of multimedia, graphics, virtual reality, amusement, design, and manufacturing. A huge number of publicly usable models such as in Google 3D Warehouse has been widely distributed and quickly spread, and many researchers attempt to provide content based retrieval techniques, e.g., sketch based 3D model retrieval [1], range image based retrieval [2], example shape based retrieval [3], and partial shape based retrieval [4], for accurately searching desirable objects and reusing these models.

A variety of 3D shape retrieval algorithms have been proposed, where early research on retrieval methods [5] focused mainly on global descriptors and their invariance under global Euclidean transformations. Recently significant effort has been invested on 3D interest point detection [6–8], local point description and organization [9], topological structure [10], non-rigid shape feature [11], and appearance analysis [12,13].

In this paper, we propose a 3D object retrieval approach by effectively utilizing low-level patches of 3D shapes consistent with geometric criterion, which are analogous to superpixels in images. In the novel framework, 3D shape is first over-segmented into many low-level patches, and different types of geometric features are extracted from these patches. Then we encode a large number of patches collected over a 3D model dataset via locality-constrained sparse coding, and extract compact and representative patch words. Input query is compared with 3D models in the database by probability distribution of patch words. Several groups of experimental results show that this method improves the retrieval performance of state-of-the-art methods.

The main contributions of the paper are described as follows.

1. Compared with point descriptors based retrieval, we introduce low-level patches to represent a 3D object, and each object only requires a small number of patches to discriminate from irrelevant objects. Moreover, these patches are not randomly generated but according to geometric criterion.
2. Different from retrieval methods based on meaningful segments and graph structure, our method avoids directly generating a few meaningful parts, which possibly leads to mis-segmented parts because the techniques of semantic segmentation are not mature. Moreover, we do not adopt graph structure to organize these parts because many topologically variable objects exist, for example, vases with different number of handles. Patch based representation will make retrieval robust against topology variation.
3. The motivation of proposing sparse coding to represent 3D objects is based on our observation, that not only the same category of objects but also irrelevant objects have many visually similar patches. For example, human body has many similar patches as horse body. Therefore, we extract sparse patch words to approximate these shapes and model occurrence frequency of these common features using locality-constrained and relatively sparser coefficients than previous bag-of-features retrieval algorithms.

* Corresponding author. Tel.: +86 29 88492344.
E-mail address: bushuhui@nwpu.edu.cn (S. Bu).

http://dx.doi.org/10.1016/j.neucom.2014.06.090
0925-2312/

The rest of this paper is organized as follows. We first discuss related works in Sections 2 and 3 present an overview of the proposed method. Section 4 introduces how to generate low-level patches and a patch segmentation algorithm. Section 5 describes each patch via different types of 3D local descriptors. The process of encoding these patches via locality-constrained sparse coding is presented in detail in Section 6. Experimental results of 3D shape retrieval are reported on two representative 3D shape data sets in Section 7, and we conclude this work in Section 8.

2. Related works

In the past decade, 3D model retrieval has become an active research topic in various fields such as multimedia, computer graphics, computer vision, and computer aided design. In contrast with early works [5] which take into account only the whole shape characteristics of 3D model, recent works mainly focus on local point description and organization, topological structure, non-rigid shape feature, and appearance analysis. Local point descriptor can encode rich local context while keeping rotation or bending invariant. Local descriptor based retrieval methods are commonly composed of three steps. (1) Detecting salient points (or local regions). This step can be omitted if all the points are uniformly sampled on the surface of 3D model. (2) Describing each salient point (or local region) with one feature vector. If this step is removed, the retrieval problem is converted into shape registration, which can be solved by Iterative Closet Points (ICP) and its invariants. (3) Comparing two sets of point (region) features. A simple way is to sum the distance of each pair of nearest neighbors in feature space. It is also feasible to compare histograms of these original features, or match template features of a training set. Representative local descriptors applied in 3D shape retrieval include global point signature [14], Laplace–Beltrami operator defined on local manifold [15], heat kernel signature [16] and scale invariant heat kernel signature [17], 3D SURF [18], 3D Harris [19], 3D SIFT [20], inner distances [21], and 3D intrinsic shape context [22]. To avoid high computational burden from combinatorial comparison between two sets of dense points, and efficiently organize these descriptors, bag of features has been borrowed from text and image processing to address correspondence and matching of 3D local descriptors [23,24].

Retrieval based on topological structure assumes that a 3D model is represented by means of a topologically connected graph consisting of nodes and edges. The problem of 3D model comparison can be converted into low-dimensional graph matching after extracting topological structures such as medial surface, curve skeleton, Reeb graph, and model graph. These simplified graphs have similar topology structures as original 3D models, and comparison between 3D models is able to be realized by virtue of checking isomorphism of simplified graphs. For example, there exist several types of graph isomorphism algorithms including tree search based algorithms, decision tree based techniques, and spectral methods. These algorithms can perform inexact computation with matching cost to measure the similarity of two simplified graphs. This type of strategies has been extensively investigated in 3D shape retrieval, for example, common undirected graph adopted in [25], Reeb graph used in [26,10] and extended Reeb graph [27], bipartite graph used in [28], skeleton graph adopted in [29] and binary tree used in [30]. Each explicitly or implicitly segmented meaningful part is identified by a single node, and edges in the graph represent adjacency relations between these segments. Therefore, shape retrieval is easily achieved by resorting to checking graph isomorphism and measuring similarity of simplified graphs of two shapes.

Many works try to solve the difficulty of non-rigid 3D shape retrieval [11] by means of utilizing various isometry invariant attributes. Geodesic distance measures the intrinsic distance between two arbitrary points on 3D surface, and contains rich geometric information. For example, geodesic distance commonly keeps unchanged under isometric deformations, which can assist in handling non-rigid shape deformation. A representative work is spectral method based on spectral decomposition of geodesic distances [31]. It filters geodesic distances appropriately to remove the effect of scaling and then compute a low-dimensional spectral embedding of 3D shape to obtain invariance to bending and rigid-body transformations. Spectral decomposition of an affinity matrix between geodesic distances characterizes a whole 3D shape, and its real eigenvalues are adopted to compare with other shapes. Diffusion distance based descriptor [32] inherits the isometry invariant attribute of geodesic distance, and further introduces the average of all the paths of fixed steps connecting two points on the surface, which is more robust than single geodesic distance. The diffusion distance is seen as average probability of traveling between two arbitrary points. Another isometry invariant descriptor is based on Laplace–Beltrami spectrum of 3D surface [33]. The spectrum is represented with eigenvalues of Laplace-Beltrami operator, and independent of different parametrization and spatial position of 3D shape. Additionally, the eigenvalues can be normalized so as to indirectly handle different scales of 3D shapes.

Appearance based 3D object retrieval tends to address how to effectively generate, organize, and compare many views of a 3D object. For example, several works focus on selecting query views [34], weighted bipartite graph matching of views [35], camera constrained-free view generation [12], constructing multiple hypergraphs of views [13], and panoramic views [36]. An interesting view based work employs interactive learning mechanism [37], which establishes a mapping from feature points in low-level feature space to points in high-level semantic space. The mechanism receives long-term relevance feedback from users via recorded retrieval history, and captures users' semantic information to refine retrieval results.

Different from previous works, in this paper we introduce the concept of low-level patches to represent a 3D object. This way avoids semantic segmentation used in retrieval methods, which is heavily dependent on topological structure such as skeleton, and unstable in the case of topological change. Moreover, the technique of locality-constrained sparse coding avails to extract high-level patch words from a large set of 3D shapes with many similar patches, which will be more compact and representative than low-level patches.

3. Overview of the proposed method

The overview of the proposed method is shown in Fig. 1. Each 3D shape is first over-segmented into a number of low-level geometric patches, and these patches are described with different types of geometric descriptors, which are adopted to characterize different geometric attributes including local, global, and topological features. After extracting these features for each patch collected from all 3D shapes in a large data set, locality-constrained sparse coding is adopted to construct a set of bases also known as visual words in a vocabulary in the domain of computer vision. Each patch is encoded by means of these bases, which are named as patch words in this paper. A number of patch words generated from a large set of 3D shapes compose a large vocabulary. Given a new object as query shape, the problem of representing it with high-level patch words is converted to optimize its coefficients via locality-constrained sparse coding.