Neurocomputing 151 (2015) 593–602

Contents lists available at ScienceDirect

Neurocomputing

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

A 3D model recognition mechanism based on deep Boltzmann machines

Biao Leng [a,*], Xiangyang Zhang [a], Ming Yao [b], Zhang Xiong [a]

[a] School of Computer Science & Engineering, Beihang University, Beijing 100191, PR China
[b] School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:
Received 14 November 2013
Received in revised form
20 February 2014
Accepted 1 June 2014
Available online 22 October 2014

Keywords:
3D model recognition
Deep Boltzmann machines
Semi-supervised learning

ABSTRACT

The effectiveness of 3D model recognition generally depends on the feature representations and classification methods. Previous algorithms have not shown good capacities to detect 3D model's feature, thus, they seem not to be competent to recognize 3D model. Meanwhile, recent efforts have illustrated that Deep Boltzmann Machines (DBM) have great power to approximate the distributions of input data, and can archive state-of-the-arts results. In this paper, we propose a novel 3D model recognition mechanism based on DBM, which can be divided into two parts: one is feature detecting based on DBM, and the other is classification based on semi-supervised learning method. During the first part, the high-level abstraction representation can be obtained from a well-trained DBM, and the feature is used in semi-supervised classification method in the second part. The experiments are conducted on publicly available 3D model data sets: Princeton Shape Benchmark (PSB), SHREC'09 and National Taiwan University (NTU). The proposed method is compared with several state-of-the-art methods in terms of several popular evaluation criteria, and the experimental results show better performance of the proposed model.

1. Introduction

The advances in computing techniques, graphics hardware, and networks have witnessed the wide applications of 3D information in various domains, such as 3D graphics, entertainment and medical industry. There is an increasing demand for effectively 3D model understanding due to the increasing number of 3D objects. Therefore, high-effective technologies to deal with 3D models are becoming highly desired [1].

In the last few decades, a great number of model descriptors have been proposed for 3D model retrieval, including statistics-based feature descriptors [2], topological representation [3], geometric-based feature descriptors [4–6] and visual representation [7–9]. All these methods just capture some low-level features, and the retrieval performance is not satisfactory. Then, hybrid features are introduced to combine different feature descriptors, such as the compact multi-view descriptor (CMVD) [10]. The features obtained from these methods are usually not expressive or discriminative enough to represent the original objects. Furthermore, they are not able to acquire the high-level semantic information of objects. Recently, extensive research efforts have been dedicated to the technique of relevance feedback [11], which can narrow the gap between high-level semantic knowledge and low-level object representation and improve the overall retrieval performance.

Deep hierarchy models have the potential to well represent the high-level abstract feature of 3D objects. The motivation of applying Deep Boltzmann Machines (DBM) [12] in 3D model recognition is inspired by the great effectiveness of deep learning. Recent research efforts related to deep learning have boomed up in many fields like speech recognition [13], natural language processing [14] and 2D image operating [15]. It has been proved that deep architectures can learn abstract features of input data. As a demonstration, the retinae system for vision of some animal is found to be a complex deep architecture [16]. The deep architecture models have powerful representative capacity, such as Restricted Boltzmann Machines (RBM), Deep Belief Networks [17], and DBM .

It is difficult to have a data set with huge numbers of labeled models because of the burdensome work to label. Therefore, extensive research has been attracted to acquire semantic information with few labels. Because obtaining labeled data for machine learning is quite difficult, while unlabeled data is usually easily accessed. Reference [18] proposes a self-taught learning method, which makes full use of unlabeled data to improve performance on supervised learning task.

In this paper, we propose a novel 3D object recognition mechanism based on DBM. DBM is firstly applied to extract the

* Corresponding author.
E-mail address: lengbiao@buaa.edu.cn (B. Leng).

http://dx.doi.org/10.1016/j.neucom.2014.06.084
0925-2312/

features from projected 2D views of 3D models. We employ a greedy layer-wise manner to train DBM, which has been demonstrated experimentally to be a practical strategy [19,20]. The DBM is expected to learn more and more abstract feature, because the feature in the top layer can be learned from the lower layers. During the training process, the contrastive divergence method is applied, aiming to push down the energy of data points. The output of last layer in DBM can be regarded as a coefficient vector of some abstract components and we take this vector as the representation of a view in a 3D model. Due to the limited scale of labeled data, a semi-supervised method is introduced to train the labeled and unlabeled data with the features obtained from the trained DBM. The experiments are conducted on the publicly available 3D object datasets: Princeton Shape Benchmark (PSB), National Taiwan University (NTU) and SHREC '09 with a few criteria. Compared with some state-of-the-art algorithms, the proposed method has better performance.

The rest of this paper is organized as following. First, we briefly introduce the related work in Section 2. Then, the proposed model for feature extraction and semi-supervised learning is presented in Section 3. Section 4 shows the experimental results in detail. Finally, we conclude this paper in Section 5.

2. Related work

During the last few decades, the field of 3D model understanding has attracted great attentions from a variety of communities, and a lot of related methods [21–28] have been proposed. Generally speaking, the 3D model understanding techniques can be divided into model descriptors and semantic understanding.

2.1. Model descriptors

In the framework of 3D model retrieval, the feature descriptor is an essential and crucial part. Due to various characteristics of 3D objects, diverse descriptors are invented to capture different features.

The statistics-based feature descriptors [29,2,6] sample 3D models and utilize the histogram as a tool for storing and comparing characteristics. Osada et al. [29] proposed a novel method called shape distribution to describe the signature of 3D models. Park et al. [2] developed a sliced image histogram to represent PCA normalized 3D objects.

Topological representation [30,31,3,32] meets the requirement of nonrigid models, such as the human body in various forms. Biasotti et al. [30] utilized the concept of size function from 2D image retrieval to reinforce the skeleton graph of a 3D model for an easier and more efficient similarity search. Patane et al. [3] presented a minimal contouring algorithm to rapidly compute the Reeb graph.

The geometric-based feature descriptors [33,4,5,34–38] contain a large group of methods, ranging from the instinctive extraction of surface characteristics to more abstract mathematical representations. Paquet and Rionx [33] firstly proposed the surface-based feature extraction method, in which the surface of a 3D model was decomposed into triangles and the normal of each was selected as a local attribute. Daras et al. [5] applied voxelization for 3D models before extracting radial and spherical information via a radial integration transform (RIT) and a spherical integration transform (SIT). Spherical-based feature descriptors had been widely applied to 3D model retrieval because of their robustness, rotational invariance and computational efficiency. Funkhouser et al. [4] proposed the spherical harmonics descriptor, where the voxel grid was treated as a binary real-valued function decomposed onto its spherical harmonics. Finally, different signatures over different radii constitute characteristics for 3D objects. Papadakis et al. [36] proposed the novel spherical functions, representing not only the intersection points of the model's surface with rays emanating from the origin but also the whole points in the direction of each ray that were closer to the origin than the furthest intersection points.

The main concept of visual representation [7,8,39–45] in 3D model retrieval is to first convert the 3D object into 2D projection images, and then utilize mature image processing techniques to extract various features. Since this approach appears quite natural to human vision, it attracts a great deal of interest. The original work was accomplished by Chen et al. [7], who proposed the light field descriptor (LFD). Papadakis et al. [45] proposed a 3D shape descriptor called PANORAMA, standing for PANormic object representation for accurate model attribution. Passalis et al. [39] suggested a novel depth buffer based descriptor called PTK. Shih et al. [40] proposed an elevation descriptor based on gray scale projection images. Since the camera view conveys important information in the area of visual representation, the number of selected views should be flexible according to various requirements. In [8], adaptive views clustering (AVC) selected the best characteristic views from more than 320 projected views. The K-means derived method and a Bayesian information criteria were employed to evaluate how likely the characteristic views fit the data. Gao et al. [9] adopted a probabilistic graph model, in which they divided the captured camera views into several sets, modeling them as a first order Markov Chain. The query comparison was converted to probabilistic analysis, which significantly alleviates the effects of transformation.

Besides global feature for projection images, some research focuses on local feature extraction [46–52]. Among them, the most fundamental work is the scale invariant feature transform (SIFT) [53]. This method has been modified and improved gradually in visual representation. Osada et al. [46] extended SIFT to Bag-of-features SIFT (BF-SIFT) and Individual match SIFT (IM-SIFT). Gao et al. [49] presented a similar algorithm called Bag-of-Region-Words (BoRW), mapping regions of views to visual words encoded from local SIFT features. Then these regions are clustered with suitable weights to represent the whole model in a set of BoRW.

2.2. Semantic understanding

The major efforts in the field of content-based 3D model retrieval concentrate on low-level features, while omitting the importance of high-level semantic information about 3D objects. Since the similarity between objects depends on subjective human observation and judgement, the need for semantic information and users' interaction are inevitable in the retrieval process. Therefore, the basic idea of semantic understanding lies in the bridging of gaps between high-level semantic information and low-level model features.

The first semantic retrieval strategy was provided by Elad et al. [54], who designed an iterative and interactive algorithm. Since then, a great deal of related work [55–57,11,58–62] has been presented. Atmosukarto et al. [55] proposed a novel relevance feedback algorithm, adopting pairwise rankings of objects. Leng and Qin [11] developed a relevance feedback concentric retrieval system, and this method dynamically updated the weights of each descriptor for improvements learning from labeled models. Ohbuchi et al. [57] developed a semi-supervised dimension reduction (SSDR) method to solve the small size sample problem in semantic retrieval.

Annotation is a way of encoding semantic features in the form of textual messages or tags, which benefits of high processing speed with respect to semantic vectors. Goldfeder and Allen [63] developed a method for the propagation of annotations attached to 3D models from Google 3DWarehouse. Later, Goldfeder et al.