Neurocomputing 151 (2015) 620–625

Contents lists available at ScienceDirect

Neurocomputing

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

3D Model Retrieval with Weighted Locality-constrained Group Sparse Coding

Xiangyu Wang [a], Weizhi Nie [b,*]

[a] Institute of Infocomm Research, Singapore
[b] School of Electronic Information Engineering, Tianjin University, China

ARTICLE INFO

Article history:
Received 15 November 2013
Received in revised form
8 February 2014
Accepted 29 March 2014
Available online 27 October 2014

Keywords:
Locality-constrained
Group sparse coding
3D model retrieval

ABSTRACT

In recent years, we have witnessed a flourishing of 3D object modelling. Efficient and effective 3D model retrieval algorithms are high desired and attracted intensive research attentions. In this work, we propose a view-based 3D model retrieval algorithm based on weighted locality-constrained group sparse coding. Representative views are first selected by clustering and the corresponding weights are provided by considering the relationship among these views. By grouping the views from 3D models, a locality-constrained group sparse coding method is employed to find the reconstruction residual for each query view. The distance between query model and candidate model is taken as the weighted sum of residual. The query model is matched to the model which can best reconstruct the query model. Experimental comparisons have been conducted on the ETH 3D model dataset, and the results have demonstrated the effectiveness of the proposed method.

1. Introduction

With the rapid development of computer hardware and software technology, more and more equipment for capturing 3D information such as Kinect has appeared in our daily life, and large amounts of 3D models are being available for computer-aided design, computer vision, etc. The fast growth of 3D model archives necessitates the progress of 3D model retrieval techniques. Effective and efficient 3D model retrieval algorithms are thus critical for many applications. The purpose of 3D model retrieval is to design algorithms computing similarities between the query and 3D models in the dataset efficiently and effectively [1]. 3D model retrieval has received great attentions in recent years. Many effective algorithms have been proposed [2–4].

The 3D model retrieval algorithms are generally divided into two categories: model-based methods and view-based methods [5]. Early works are mainly model-based methods, in which low-level features or high-level structure-based information are employed. They usually focus on 3D model representation. Features such as geometric moment [2], surface distribution [6], volumetric descriptor [7], and surface geometry [8] are usually utilized. Due to the requirement of 3D models, these methods are limited in the practical applications.

View-based methods are becoming popular and extensive efforts have been dedicated to this area, because of the high discriminative property of multi-views for 3D object representation [9]. View-based methods represent 3D models using a set of 2D images. The 2D image sets show the 3D objects from different views, which contain both the spatial and structure information. The view-based method has two advantages:

- First, it does not require the explicit virtual model information.
- Second, visual analysis also plays an important role. The view-based 3D model analysis methods can benefit from the existing image processing technologies.

Many view-based 3D object retrieval methods, such as LightField Descriptor (LFD) [3], Elevation Descriptor (ED) [10], Bag-of-Visual-Features (BoVF) [9], SSCD [4] have been proposed these years. For 3D model matching, the 3D model comparison is based on the comparison of features from different views. In [11], Zernike moments and Fourier descriptors are employed to describe each 2D image. Recently, Gao et al. [4] proposed a spatial structure circular descriptor (SSCD) method to compare two 3D models. The SSCD can preserve the global spatial structure of 3D models, and it was invariant to rotation and scaling. All spatial information of 3D model can be represented by an SSCD which included several SSCD images. In this method, the 3D model is projected to a plane, and the comparison between two 3D models is based on the matching between two groups of SSCD images. However, SSCD needs a fully 3D model in order to generate the 2D image in different planes, which limits its applications.

* Corresponding author.
E-mail address: truman.nie@gmail.com (W. Nie).

http://dx.doi.org/10.1016/j.neucom.2014.03.091
0925-2312/

There are two main stages for 3D model retrieval, including 3D model representation and 3D model matching [12]. Most of existing works focused on 3D model representation methods. For 3D model matching, the comparison between two groups of 2D views is modelled as a weighted bipartite graph matching (WBGM) in [12]. The weighted bipartite graph is built with the selected representative 2D views, and the proportional max-weighted bipartite matching method is employed to find the best match in the weighted bipartite graph. However, the relationship among views from different models are neglected. Other methods such as [13] involve a learning process. 3D model retrieval can be considered to find the 3D model which can reconstruct the query views most accurately. Thus, it can be modelled as a reconstruction task based on sparse coding. In this work, we propose a view-based 3D model retrieval algorithm based on weighted locality-constrained group sparse coding. In the view-based 3D model retrieval algorithm, we need to match query model with models in the dataset. Each model is represented by a set of 2D views. For the query model, representative views are first selected and the corresponding weights are provided by considering the relationship among these views. The query model is matched to the model which can best reconstruct the query model. By grouping the views from 3D models, a locality-constrained group sparse coding method is employed [14] to find the reconstruction residual for each query view. The distance between query model and candidate model is taken as the weighted sum of residual.

The remainder of this paper is organized as follows. Section 2 briefly reviews related work. The proposed 3D model retrieval algorithm using weighted group sparse coding method is presented in Section 3. Experimental results and discussions are provided in Section 4. Finally, we conclude the paper in Section 5.

2. Related work

In this section, we focus on view-based 3D object retrieval works. In the past few decades, content-based image retrieval (CBIR) that finds relevant images from a database by providing one or multiple visual examples gained a lot of research efforts [15]. View-based 3D object retrieval is a special CBIR task as each 3D object is represented by a set of 2D view images that are usually captured from a group of cameras . Typically, there are two key issues in the technology, i.e., 3D model representation and 3D model matching. Most of existing works focused on 3D model representation methods. Different features are proposed to extract information from 2D view images. LightField Descriptor (LFD) [3] is one famous view based shape descriptor. In LFD, each group of representative views includes 10 views. One hundred orthogonal projections of an object, excluding symmetry, are encoded both by Zernike moments and Fourier descriptors as features. In Compact Multi-View Descriptor (CMVD) [16], a set of 2D rotation-invariant shape descriptors, based on the Polar-Fourier Transform, Zernike Moments and Krawtchouk Moments, is produced as the final set of descriptor vectors. Recently, Gao et al. proposed another view based descriptor named spatial structure circular descriptor (SSCD) in [4]. It can preserve the global spatial structure of 3D models, and is invariant to rotation and scaling. The comparison scheme for these approaches highly depends on the view generation methods, which limits the application of this type of view-based 3D retrieval methods.

There are also some works focusing on the 3D model matching. Adaptive Views Clustering (AVC) [17] first selects representative views from the initial captured views. A Bayesian probabilistic method is then employed to retrieve similar 3D objects. The authors proposed a Hausdorff distance learning method in [18], which can automatically learn the optimal distance metric for 3D object matching. Weighted bipartite graph matching has been employed in [12] to find the matched 3D object for the query. Two sets of views from query and model are modelled as a bipartite graph, and the weights are updated with random walk. Finally, the weighted bipartite graph matching is conducted to compute the distance between 3D objects. In [13], a hypergraph is constructed to formulate the relation among 3D objects by view clustering. The semi-supervised learning on the hypergraph is conducted to estimate the object relevance for 3D object retrieval. It can be observed that most of the methods focus on the distance estimation between objects. However, they fail to capture the structure information of multiple views as well as the correlation clues among different objects. Different from other view selection methods, which are based on clustering or pre-defined camera array, Gao et al. [19] proposed to interactively select representative views with user relevance feedback. This method is able to find the most discriminative view for the current query and also learn optimal distance metric, which can achieve satisfied retrieval results.

In this work, we propose a weighted locality-constrained group sparse coding based approach to obtain the group based representation of query object and then combine the reconstruction residuals according to the weights of query views. Objects from the same category are taken as a group. By employing a locality-constrained group sparse coding method, it captures the inter-group and intra-group relation to some extent.

3. Proposed method

In this section, we introduce the proposed 3D model retrieval using weighted locality-constrained group sparse coding method. First the scheme of our approach is described, and following the detail algorithm is given.

3.1. Overview of the framework

Given the query model Q, the goal of 3D model retrieval is to find the matched models from the 3D model dataset $\mathcal{M} = \{M_1, \ldots, M_K\}$. Let $\mathcal{V}^Q = \{\mathbf{v}_1^Q, \ldots, \mathbf{v}_m^Q\}$ denote the view set of the query model Q with m views, and Let $\mathcal{V}^M = \{\mathbf{v}_1^M, \ldots, \mathbf{v}_n^M\}$ denote the view set of model M in the 3D model dataset with n views.

The key topic is to compute the distance between query object and 3D models. In order to represent each view in the 3D model, various features can be used. In this work, Zernike moments descriptor is employed as the feature for each 3D model. Zernike moments have been demonstrated to be robust to image translation, scaling, and rotation, and applied in many 3D model analysis tasks. For each 3D model, the Zernike moments descriptor is extracted from each 2D view, which captures the spatial and structure information of 3D model. Each view is represented by a 49-D Zernike moment feature vector. For the query model Q, the corresponding feature vector of $\mathbf{v}_i^Q$ is denoted as $\mathbf{f}_i^Q$. For model M in the 3D model dataset $\mathcal{M}$, the corresponding feature vector of $\mathbf{v}_i^M$ is denoted as $\mathbf{f}_i^M$.

For the query object, the representative views are first selected from the original view set by clustering similar views. By jointly considering the size of clusters and visual similarity among selected views, the corresponding weights are obtained using random walk method. For 3D models in the dataset, the views from models are clustered into G groups based on their category. The obtained features $X = \{X_1, \ldots, X_G\}$ where X_i corresponds to the features in ith group are taken as the dictionary. The problem of finding the matching view for query model is formulated as the reconstruction process for the query view with the dictionary X. The Group Lasso [20] is employed. For each model, the weighted