Signal Processing 120 (2016) 702–713

Contents lists available at ScienceDirect

Signal Processing

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

Fast view-based 3D model retrieval via unsupervised multiple feature fusion and online projection learning

Jun Xiao, Yinfu Feng*, Mingming Ji, Yueting Zhuang

School of Computer Science, Zhejiang University, Hangzhou 310027, PR China

ARTICLE INFO

Article history:
Received 5 September 2014
Received in revised form
21 November 2014
Accepted 27 November 2014
Available online 5 December 2014

Keywords:
View-based 3D model retrieval
Multiple feature fusion
Online projection learning
Out-of-sample

ABSTRACT

Since each visual feature only reflects a unique characteristic about a 3-dimensional (3D) model and different visual features have diverse discriminative power in model representation, it would be beneficial to fuse multiple visual features in 3D model retrieval. To this end, we propose a fast view-based 3D model retrieval framework in this article. This framework comprises two parts: the first one is an Unsupervised Multiple Feature Fusion algorithm (UMFF), which is used to learn a compact yet discriminative feature representation from the original multiple visual features; and the second one is an efficient Online Projection Learning algorithm (OPL), which is designed to fast transfer the input multiple visual features of a newcome model into its corresponding low-dimensional feature representation. In this framework, many existing ranking algorithms such as the simple distance-based ranking method can be directly adopted for sorting all 3D models in the database using the learned new feature representation and returning the top ranked models to the user. Extensive experiments on two public 3D model databases demonstrate the efficiency and the effectiveness of the proposed approach over its competitors. The proposed framework cannot only dramatically improve the retrieval performance but also reduce the computational cost in dealing with the newcome models.

1. Introduction

With the rapid development of computer hardware and computer graphics techniques, especially the modeling and rendering techniques, more and more 3-dimensional (3D) models have been created and used in a wide range of applications [1,2]. Several public large-scale 3D model databases [3,4] are also available on the internet. To manage and reuse the abundant 3D models, efficient and effective 3D model retrieval techniques and systems become crucially important.

In some early researches, the keyword-based methods were applied to retrieve 3D models whose tags are similar to the query from database. These methods require the user to label all 3D models in the database in advance, which is very tedious and time-consuming. Moreover, due to the high ambiguity and subjectivity of keyword, it is difficult to choose a correct and meaningful keyword as the tag for a given 3D model. It is also hard to measure the similarity between two visual or geometric similar 3D models with different keywords. So, these methods are inefficient and of low accuracy in many cases.

To overcome the shortcomings of keyword-based methods, a growing number of researchers have focused their attentions on the content-based 3D model retrieval techniques [5–8]. In [9], the authors have shown that the content-based methods consistently outperform the keyword-based

* Corresponding author.
E-mail addresses: junx@zjuem.zju.edu.cn (J. Xiao), yinfufeng@zju.edu.com (Y. Feng), mjcs@zju.edu.cn (M. Ji), yzhuang@zju.edu.cn (Y. Zhuang).

http://dx.doi.org/10.1016/j.sigpro.2014.11.020
0165-1684/

methods. Since efficient feature representation is the cornerstone of these content-based methods, a great deal of effort has been devoted to extracting or constructing various visual and geometric features including the low-level features (e.g., geometric moment [10] and surface features [11]) and the high-level structure features (e.g., MATE feature [12]) from 3D models as feature representation. However, many existing methods tended to use only one type of visual or geometric feature representation to characterize the 3D model, which results in poor performance. Recently, more and more researchers have found that correctly combining multiple features can significantly improve the algorithms' performance in a large variety of areas [13–19]. In these work, how to combine/fuse multiple features becomes an important problem. An intuitive way is to concatenate multiple features into a new high-dimensional feature vector, which is very simple while lacks of physical meaning. It would be more smart to take both the complementary information between different feature representation and the intrinsic geometric structure information of each feature representation into account in fusing these multiple features. More importantly, how to deal with the newcome samples, which is named as the out-of-sample problem, is another open problem. However, not much work has been done to solve these problems well, especially for the task of 3D model retrieval.

To address the above mentioned issues, we propose a fast view-based 3D model retrieval framework. The framework comprises two parts: one is an Unsupervised Multiple Feature Fusion algorithm (UMFF) and the other one is an efficient Online Projection Learning algorithm (OPL). We first use UMFF to learn a more compact yet discriminative feature representation from the original multiple features. Then, OPL is designed to reduce the high computational cost in solving the objective function of UMFF for the newcome models by learning a projection matrix. Once the projection matrix is obtained, the input multiple high-dimensional features of the newcome 3D model can be fast transferred into the new compact yet discriminative feature representation. So the out-of-sample problem can be overcome via OPL in this work. In the proposed framework, many ranking algorithms such as the distance-based ranking method can be directly used to sort the 3D models and return the top ranked models to the user.

The rest of this article is organized as follows. Section 2 reviews some related work on the topic of 3D model retrieval. Then, the details of our proposed framework is introduced in Section 3. Section 4 presents the experimental results on two public 3D model databases. Finally, Section 5 concludes our work and sketches several direction of future work.

2. Related work

The content-based 3D model retrieval methods can be roughly divided into two categories: 3D model based and view based [8,20]. In 3D model based methods, the model representation includes the low-level features like the geometric moment [10] and surface distribution [11], and the high-level structure-based features [12]. All of these feature representations are directly extracted from the 3D model data. This kind of methods requires less computational cost for feature extraction, but it needs to access the geometric information about the model, which is infeasible in some cases [21–24]. In contrast, view-based 3D model retrieval methods use the projection images of the 3D model from multiple views as inputs, which can provide as much information as the model itself, while they don't need to access the geometric information about the model. Thanks to the discriminative property of multiple views of 3D model, the view-based methods have received encouraging performance in the task of 3D model retrieval in recent years. However, the great success of the view-based methods heavily rely on two factors: (1) the feature representation of the projection images; (2) how to combine/fuse these multi-view based features.

In the work [25] Shih et al. proposed an Elevation Descriptor (ED), which utilizes the altitude information of the 3D model from six directions for feature representation. Ohbuchi et al. [26] used a 2D Fourier descriptor to characterize the depth or z-value images of a 3D model for 3D model similarity matching. In [27], the shape descriptor based on Curvature Scale Space (CSS) representation of the contour was extracted from 7 characteristic views for multiple views matching. Ansary et al. [28] selected 49 coefficients of the Zernike moment descriptor to represent every 2D view-image in their method. However, these methods used only one kind of visual feature for describing each view of the model, which leads to the weak representation power of their feature.

Meanwhile, multiple features have been widely used in computer vision and multimedia communities for boosting the algorithms' performance in various applications [13,14,16]. In light of this, more and more recent researches on 3D model retrieval have tried to extract multiple features as 3D model feature representation. Chen et al. [5] proposed a Light Field Descriptor (LFD), which generates 10 representative views from the vertices of a dodecahedron as 3D model representation. In LFD, Zernike moments and Fourier descriptors were employed as the feature representation to describe each representative view. Similarly, the Compact Multi-View Descriptor (CMVD) [14] extracted Polar-Fourier Transform, Zernike Moments and Krawtchouk Moments to describe the multi-view images which are generated through 18 vertices of the bounding 32-hedron of each 3D model. Li and Johan [29] used a hybrid features named ZFDR to describe a 3D model, which is linearly combined Zernike moments, Fourier descriptors, the depth information and the Ray-based features together. Ohbuchi and Furuya [30] first extracted 3 kinds of visual features using the Scale Invariant Feature Transform (SIFT) [31] algorithm as the interest point detector. Then an unsupervised distance metric learning method based on Manifold Ranking (MR) [32] was implemented to compute distances between these features in their work. To combine these multiple features, they employed a linear combination of the distances derived from these multiple features. In summary, multiple features in the above mentioned methods were just simply combined together with a linear model or treated independently. Therefore, they are failed to exploit the complementary information between different feature representations.