J. Vis. Commun. Image R. 37 (2016) 40–45

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

3D object retrieval based on sparse coding in weak supervision ☆

Wei-Zhi Nie, An-An Liu *, Yu-Ting Su

School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 14 December 2014
Accepted 7 June 2015
Available online 25 June 2015

Keywords:
3D model retrieval
Sparse representation
Dictionary learning
Fisher discrimination
Weak supervision
Characteristic view extraction
Similarity measure
View-based model

ABSTRACT

With the rapid development of computer vision and digital capture equipment, we can easily record the 3D information of objects. In the recent years, more and more 3D data are generated, which makes it desirable to develop effective 3D retrieval algorithms. In this paper, we apply the sparse coding method in a weakly supervision manner to address 3D model retrieval. First, each 3D object, which is represented by a set of 2D images, is used to learn dictionary. Then, sparse coding is used to compute the reconstruction residual for each query object. Finally, the residual between the query model and the candidate model is used for 3D model retrieval. In the experiment, ETH, NTU and ALOL dataset are used to evaluate the performance of the proposed method. The results demonstrate the superiority of the proposed method.

1. Introduction

With the development of computer vision technologies, more and more 3D data collected equipment have appeared in our life and lead to huge amount of 3D models. 3D models have been widely used in many applications such as computer graphics, computer-aided design, computer vision, and medical image processing. Compared to 2D images, 3D models can convey more information about the object. As shown in Fig. 1, instead of showing only one view, the 3D image describes multiple views of the object. The fast growth of 3D model achieves necessitate the progress of 3D model retrieval techniques. The purpose of 3D model retrieval is to design algorithms computing similarities between the query and 3D models in the dataset efficiently and effectively [3]. 3D model retrieval has attracted considerable attention. Many effective algorithms are proposed in recent years [32,4,24,20,39].

There are totally two categories of 3D model retrieval algorithms: model-based methods and view-based methods [11,13,7,44]. Early works are mainly model-based methods, in which low-level feature-based methods or high-level structure-based methods are employed. Low-level feature-based method utilized geometric moment [31], surface distribution [29], volumetric information [34] and surface geometry [30] to describe 3D model. Many low-level features are also based on the 2D image feature, and 3D model retrieval depends on the low-level feature comparison. At the same time, high-level structure requires 3D model to be available explicitly [33]. Each 3D model must have the clearly spatial and structural information, which limits the practical applications of model-based methods.

View-based method is the other category of popular 3D model retrieval algorithm recently [11,13,6,14]. View-based methods represent 3D models using a set of 2D images. This image set shows the 3D object from different views. Therefore, these 2D images have the spatial and structure information. The key point is how to extract feature from these 2D images to represent spatial and structure information for each 3D model. Zernike moments and Fourier descriptors [17] are employed to describe each 2D image. Elevation descriptor (ED) [33] has been proposed recently. Six elevations are obtained from six different views to describe 3D models. 3D model comparison is based on the comparison of these six elevations. Recently, Gao et al. [9] proposed a spatial structure circular descriptor (SSCD) method to compare two 3D models. In this method, the 3D model is projected to a plane, and the comparison between two 3D models is based on the matching between the two groups of SSCD images. However, SSCD need a fully 3D model in order to generate the 2D image in different planes. It is hard to promote in reality. This is due to the fact that view-based methods are with the highly discriminative property for object representation and visual analysis also plays an important role in multimedia applications. The view-based method has two advantages. First, it does not require the explicit virtual model information, which makes the method more applicable to real practical application.

☆ This paper has been recommended for acceptance by Dr. Ankit Chaudhary.

* Corresponding author.
E-mail address: anan0422@gmail.com (A.-A. Liu).

http://dx.doi.org/10.1016/j.jvcir.2015.06.011
1047-3203/

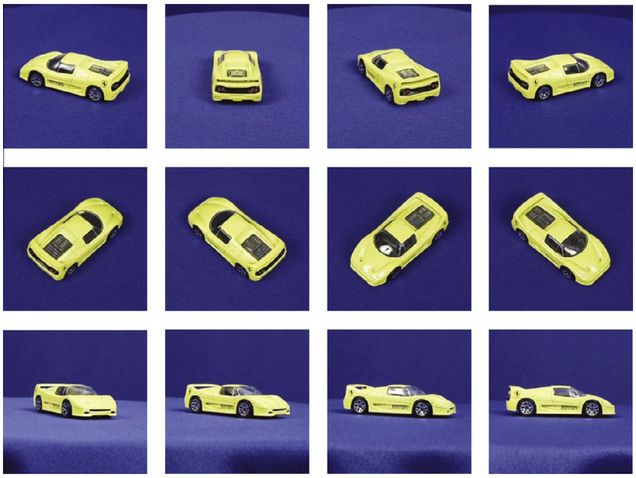

Fig. 1. One 3D object examples described by multiple views.

Second, image processing has been investigated for many decades, The view-based 3D model analysis methods can benefit from existing image processing technologies.

In this paper, we proposed a novel 3D model retrieval algorithm based on sparse representation based classification. We first extract feature from these 2D images, then apply Fisher Discrimination Dictionary Learning algorithm [36] to single dictionary for each 3D model [36]. Then, SRC [35] is used to find the reconstruction residual for each query object. Finally, 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.

The proposed scheme has three advantages in the following aspects:

- The proposed method does not require the explicit virtual model information. We extract 2D images form each 3D model, and apply the features of 2D images to represent 3D model. Thus, some traditional image progress technologies can be utilized in this work.
- We employ FDDL to learn structure dictionary for each object, which can well reconstruct query sample for model retrieval. FDDL method can learn structured dictionary which includes a set of sub-dictionaries. This design can represent the classification of the 3D dataset for the accurate reconstruction error.

The paper is organized as follows. Section 2 briefly reviews related work. Section 3 introduces the proposed method, which will detail some key technologies in this study. Experiments are provided in Section 4, which will demonstrate the performance of our approach. Finally, The conclusion is stated in Section 5.

2. Related work

In this section, we mainly review the related works in the view-based techniques for 3D object retrieval. Meanwhile, based on 3D model retrieval problem, each object cannot provide enough images to learn dictionary. Thus, sparse coding method can be seen as one weakly supervised classification method to learn structural dictionary. We also introduce some weakly supervised learning method in this section [37].

(1) In the view-based 3D model retrieval, features are extracted based on the 2D images. The visual similarities between the different view images from different models are compared with each other to measure the difference of the models. Light Field [5] is a famous view based shape descriptor. It defines the distance of two 3D models as the minimal distance between their 10 corresponding silhouette views. To measure the difference of two silhouette views, it uses a hybrid image metric [26] which integrates the Zernike moments descriptor and Fourier descriptor. Multiple view descriptor first aligns the model with PCA [18] and then classifies models by comparing the primary, secondary and tertiary views defined by the principle axes. Salient local visual feature-based retrieval method [27,7] first renders a set of depth view images for a 3D model and then extracts the multi-scale local features of these views using SIFT [22], which is invariant to translation, scaling and rotation. Finally, it fuses all of local feature into a histogram using BOW approach, which accumulates the visual words of multiple views into a single histogram to represent the feature of a 3D model. Recently, Gao et al. [9] proposed another view based descriptor named spatial structure circular descriptor (SSCD), which can preserve the global spatial structure of 3D models, and is invariant to rotation and scaling. In addition, all spatial information of 3D models can be represented by SSCD without redundancy. Because this method need a fully 3D model to generate the spatial structure circular descriptor, it is hard to use in real living.

(2) Weakly supervised learning methods have been used in many applications [38,43]. Zhang et al. [41] present a novel weakly supervised image segmentation algorithm by learning the distribution of spatially structured superpixel sets from image-level labels. Zhang et al. [42] also proposed a weakly supervised photo cropping, where a manifold embedding algorithm is developed to incorporate image-level semantics and image global configurations with graphlets, or small-sized connected subgraph. Oquab et al. [28] proposed a weakly supervised convolution neural network (CNN) for object classification that relies only on image-level labels, yet can learn from cluttered scenes containing multiple objects. Zhong et al. [45] consider visual tracking in a novel weakly supervised learning scenario where labels but no ground truth are provided by multiple imperfect oracles.

3. Proposed method

In this section, we introduce the proposed 3D object retrieval based on sparse coding method. First, the formulation of the proposed method is introduced, and then the details algorithm is given.

3.1. Problem formulation

The goal of retrieval is to find the best matching models form the candidate dataset $M = \{M_1, M_2, \ldots, M_c\}$. Let $v^Q = \{v_1^Q, \ldots, v_m^Q\}$ denote the view set of query object Q, which includes m views. $v^{M_c} = \{v_1^{M_c}, \ldots, v_n^{M_c}\}$ denote the view set of candidate 3D model with n views.

The main problem is to compute the distance or similarity between query model and candidate 3D models. For feature extraction, Zernike moments descriptor is utilized as the feature for each view, which captures the spatial and structure information of 3D model. Each view is represented by a 49-D Zernike moments feature vector [40]. Each view v can be denoted as f. For 3D models in the dataset, different views belong to different objects, and different objects belong to different categories. We can directly employ these original features of each view as dictionary for sparse