Neurocomputing 207 (2016) 684–692

Contents lists available at ScienceDirect

Neurocomputing

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

Multi-scale object retrieval via learning on graph from multimodal data

Yongsheng Zhang*, Tsuyoshi Yamamoto, Yoshinori Dobashi

Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan

ARTICLE INFO

Article history:
Received 10 January 2016
Received in revised form
8 May 2016
Accepted 16 May 2016
Communicated by Ke Lu
Available online 4 June 2016

Keywords:
Object retrieval
Multi-scale
Learning on graph
Multimodal data

ABSTRACT

Object retrieval has attracted much research attention in recent years. Confronting object retrieval, how to estimate the relevance among objects is a challenging task. In this paper, we focus on view-based object retrieval and propose a multi-scale object retrieval algorithm via learning on graph from multimodal data. In our work, shape features are extracted from each view of objects. The relevance among objects is formulated in a hypergraph structure, where the distance of different views in the feature space is employed to generate the connection in the hypergraph. To achieve better representation performance, we propose a multi-scale hypergraph structure to model object correlations. The learning on graph is conducted to estimate the optimal relevance among these objects, which are used for object retrieval. To evaluate the performance of the proposed method, we conduct experiments on the National Taiwan University dataset and the ETH dataset. Experimental results and comparisons with the state-of-the-art methods demonstrate the effectiveness of the proposed method.

1. Introduction

Recent progress on computer devices and graphic processing technology has lead to wide applications of 3D objects [1,2], such as industrial design, computer graphics, and virtual reality. These increasing applications generate large scale 3D object data and lead to requirements for advanced 3D data processing technique, such as 3D saliency detection [3]. How to manage such data becomes a challenging task, which requires effective and efficient object retrieval methods.

In recent years, 3D object retrieval [4] has attracted much research attention from both research and industrial fields. Extensive research attention [5–7] has been dedicated in such an emerging field [8–11], from either model-based [12–14] or view-based directions [15–18], based on the representation methods of 3D objects.

Most of early objects are model-based, where each object is represented by a virtual 3D model, such as triangle mesh. For these methods, low level features, such as volumetric descriptor [19], surface distribution [13] and geometry [12,20,21] can be extracted for object description. Other methods extract high level feature for object structure description. Model-based methods have shown advantages when describing the global spatial information. However, one main limitation of such methods lies in situation of lack of model data. Model-based methods highly depend on the virtual model, where such model information may be not available in many practical applications. Although 3D model reconstruction has been investigated for decades, it is still a challenging task and costs much computation load. Under such circumstance, model-based methods cannot work well.

Different from model-based methods, view-based methods [15,22,16,23] have become more useful in recent years. In view-based methods, each object is represented by a set of views from different directions. Such methods are much more flexible than model-based methods, as the model information is not mandatorily required. Also, view-based methods can be beneficed from existing image processing achievements, such as image feature extraction and comparison. Fig. 1 provides examples of views from objects. Daras et al. [24] introduced that view-based methods [25,26] can be more discriminative than model-based methods, and view-based methods have been investigated in recent decade.

In view-based methods, generally, object comparison is based on multi-view matching. It is noted that it is still a challenging task to compare two objects via two groups of views, which is different from traditional image comparison. In this paper, we focus on view-based object retrieval and propose a multi-scale object retrieval algorithm via learning on graph. Fig. 2 shows the framework of our proposed method. In our work, shape features are extracted from each view of objects. The relevance among objects is formulated in a hypergraph structure, where the distance of different views in the feature space is employed to generate the connection in the hypergraph. To achieve better representation performance, we propose a multi-scale hypergraph structure to

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2016.05.053
0925-2312/

Fig. 1. Example views of 3D objects.

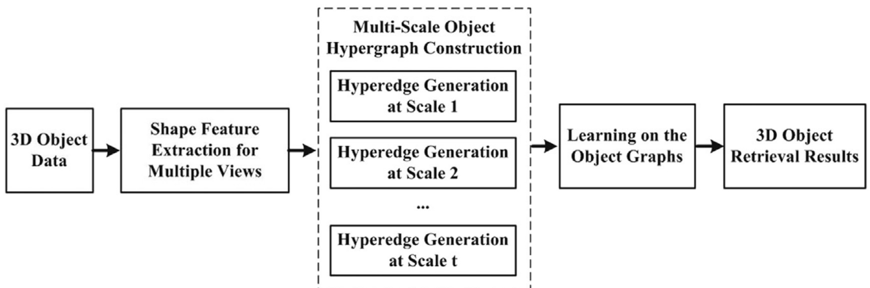

Fig. 2. The framework of the proposed method.

model object correlations. The learning on graph is conducted to estimate the optimal relevance among these objects, which are used for object retrieval. The hypergraph structure is able to exploit the high-order correlation among different data modalities, features, or even data types. In this way, the hypergraph modeling can be used to formulate the correlation among multiple views of 3D models. In our work, the representation is based on multi-scale view comparison, which leads to the view matching from different aspects. As the direct view matching is difficult, the proposed scheme is able to explore all possible view connection and yet achieves better performance. To evaluate the performance of the proposed method, we conduct experiments on the National Taiwan University dataset and the ETH dataset. Experimental results and comparisons with the state-of-the-art methods demonstrate the effectiveness of the proposed method.

The contributions of our work are two-folds.

1. We propose a multi-scale formulation from the data comparison aspect for object retrieval based on graph learning. More specifically, this work can be further extended to support more data modalities and features.
2. We conduct extensive experiments on two datasets, and the results can justify the performance of proposed method.

This paper is organized as follows. Section 2 introduces related work on 3D object retrieval. Section 3 provides the proposed method. Experimental results are provided in Section 4 and conclusions are given in Section 5.

2. Related work

In this section, we will introduce recent progress on 3D object retrieval. 3D object retrieval can be divided into two types of methods, i.e., model-based methods and view-based methods.

For model-based methods, low-level features [19,20,12] and high-level features can be employed. To extract model-based feature, Papadakis et al. [27] introduced a panoramic view, i.e., panoramic object representation for accurate model attributing (PANORAMA), which was generated by projecting the model to a lateral surface of a cylinder. To compare two 3D models, the distance was measured by matching between two PANORAMA images. In [28], Gao et al. introduced a spatial structure circular descriptor (SSCD), which employed the projected model information in a circular region to represent the 3D model. In SSCD, the projected image is able to preserve the global spatial information, and the comparison between two 3D models is achieved by the histogram distance for each SSCD view. Vranic et al. [29] introduced an Extension Ray-based Descriptor (ERD) method, where the concentric spheres were used to extract the surface information. In this method, each sampling surface point had a value on the corresponding sphere surface.

Chen et al. [30] proposed the first view-based 3D object retrieval method, i.e., Lighting Field Descriptor (LFD). In LFD, several groups of 10 views are used to represent each 3D object. For these views, the Zernike moments and Fourier descriptors were employed as the features. To compare two 3D objects, the minimal distance between two groups of views was used in [30]. Shih et al. [31] introduced an Elevation Descriptor (ED), which employed six range views from different directions to represent 3D objects. In ED, the depth histogram was extracted as the ED feature and matching between two ED histograms was measured as the distance between two 3D objects. Daras et al. [24] introduced the Compact Multi-View Descriptor (CMVD), which contained 18 views from the vertices of a 32-hedron. Mahmoudi et al. [32] proposed to employ the curvature scale space as the view descriptor. The curvature scale space was combined with Zernike moments to compare 3D models. Adan et al. [33] proposed a depth gradient image (DGI) model, which employed both the surface and the contour information to avoid restrictions concerning the layout and visibility of 3D models. Zhang et al. [34] proposed to employ bipartite graph matching for 3D model comparison. Ji et al. [35] proposed an efficient semi-supervised multiple feature fusion method for comparison. The bag of salient local spectrums was introduced in [36] for non-rigid and partial 3D object retrieval.

To select representative views from the large view pool, a query