Signal Processing 112 (2015) 110–118

Contents lists available at ScienceDirect

Signal Processing

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

View-based 3D object retrieval via multi-modal graph learning

Sicheng Zhao, Hongxun Yao*, Yanhao Zhang, Yasi Wang, Shaohui Liu

School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

ARTICLE INFO

Article history:
Received 11 March 2014
Received in revised form
25 September 2014
Accepted 27 September 2014
Available online 14 October 2014

Keywords:
3D object retrieval
Multi-modal graph learning
Feature fusion
Multi-views

ABSTRACT

Content-based 3D object retrieval has wide applications in various domains, ranging from virtual reality to computer aided design and entertainment. With the rapid development of digitizing technologies, different views of 3D objects are captured, which requires for effective and efficient view-based 3D object retrieval (V3DOR) techniques. As each object is represented by a set of multiple views, V3DOR becomes a group matching problem. Most of state-of-the-art V3DOR methods use one single feature to describe a 3D object, which is often insufficient. In this paper, we propose a feature fusion method via multi-modal graph learning for view-based 3D object retrieval. Firstly, different visual features, including 2D Zernike moments, 2D Fourier descriptor and 2D Krawtchouk moments, are extracted to describe each view of a 3D object. Then the Hausdorff distance is computed to measure the similarity between two 3D objects with multiple views. Finally we construct multiple graphs based on different features and learn the optimized weights of each graph automatically for feature fusion task. Extensive experiments are conducted on the ETH-80 dataset and the National Taiwan University 3D model dataset. The results demonstrate the superior performance of the proposed method, as compared to the state-of-the-art approaches.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

With the rapid development of digitizing technologies, the process of generating 3D models is made much easier. Compared to 2D data, 3D data has high superiority in stereo information representation and processing, which makes 3D data widely applied in various domains, ranging from virtual reality to computer aided design and entertainment [1,2]. As a result of the increasing use of 3D information acquisition equipment by the public, more and more 3D models are created and shared, which results in an urgent requirement to retrieve these data efficiently and effectively. Thus content-based 3D object retrieval (3DOR) arose and has attracted much attention of researchers.

The goal of 3DOR is to search for 3D models that are similar to the query model in content level. The main task in 3DOR is defined in [3] as follows: given a query object, define appropriate measures to automatically assess the similarity between any pair of 3D objects based on a suitable notion of similarity.

Generally, according to the different employed data type and the corresponding methods, existing 3DOR methods can be divided into two categories: model-based methods (M3DOR) and view-based methods (V3DOR). Most of the early 3D object retrieval methods [4,5] belong to M3DOR, which require explicitly virtual 3D model information, such as point cloud data. M3DOR can also be categorized into two types: geometry-based approach and visual similarity-based approach. The geometry-based approach concentrates on the shape-based [5] or topology-based matching [6], which utilizes the distribution of vertices or polygons, or the topological structures of 3D model to measure the similarity. The visual similarity-based approach uses visual features to represent 3D models, such as LightField Descriptor [7]. The

* Corresponding author. Tel.: +86 18645045973.
E-mail addresses: zsc@hit.edu.cn (S. Zhao), h.yao@hit.edu.cn (H. Yao), yhzhang@hit.edu.cn (Y. Zhang), wangyasi@hit.edu.cn (Y. Wang), shliu@hit.edu.cn (S. Liu).

http://dx.doi.org/10.1016/j.sigpro.2014.09.038
0165-1684/© 2014 Elsevier B.V. All rights reserved.

advantage of M3DOR is that all the model information can be employed, while this also leads to the major limitation of M3DOR, as 3D model information is not so easy to obtain from real objects. The limitations severely affect the practical applications of M3DOR methods.

Fortunately, with the rapid development of camera technology, we are able to capture different views of 3D objects easily. Thus view-based 3D object retrieval (V3DOR) methods, which represent 3D objects by a set of 2D images, become popular [8–11]. Compared to M3DOR, V3DOR has three advantages [10,12]: (1) V3DOR is more flexible because the virtual 3D model information is not mandatory required; (2) it is discriminative for articulated objects and effective for partial matching; it is also beneficial for 2D sketch based and 2D image based queries; (3) it can benefit from existing image processing technology for the importance of visual analysis in the methods. Experimental results in [13,14] demonstrate that V3DOR can achieve better performance than M3DOR.

There are two major steps for V3DOR: view representation and object matching. Many efforts have been devoted to the representation of 3D views, such as Compact Multi-View Descriptor [10], Elevation Descriptor [15] and Bag-of-Region-Words [8,16]. However, these approaches mainly use one single feature and the comparison scheme for them highly depends on the view generation methods. 3D object matching methods usually focus on the distance measure between different 3D objects. As each object is represented by a set of different views, the matching problem between two objects can be formulated as a group (many-to-many) matching problem. Many different matching techniques have been exploited, such as Hausdorff distance and Sum distance [15,11].

In this paper, we focus on the V3DOR method and mainly tackle the view representation problem in a feature fusion framework. The framework of the proposed method is shown in Fig. 1. Contrast to most of the state-of-the-art methods which use single feature, we adopt multiple features to enhance the representation capability, including 2D Zernike moments, 2D Fourier descriptor and 2D Krawtchouk moments. The Hausdorff distance is computed to measure the similarity between two 3D objects with multiple views. Then a feature fusion framework is proposed based on multi-modal graph learning. We construct multi-modal graphs for all features fused with different weights and learn the weights in a regularization framework in order to sufficiently explore the complementation of different features. We conduct experiments on ETH-80 and NTU 3D datasets to demonstrate the effectiveness of the proposed method. Experimental results show that the proposed method outperforms the state-of-the-art of single-graph based methods for V3DOR.

In summary, the main contributions of this paper are two-fold.

1. We test the retrieval performance of different features, which describe different 2D geometric properties of 3D objects, based on single-graph learning for V3DOR. From the results we conclude that different features perform differently on different datasets.
2. We propose to use multi-modal graph learning as a feature fusion method for V3DOR. The optimized weights of each graph are automatically learnt in a regularization framework to effectively utilize the complementary associations of different features to improve the retrieval performance. It should be noted that the proposed method can be easily generalized to other retrieval tasks.

The rest of this paper is organized as follows. Section 2 briefly reviews the related work on 3D object retrieval. Section 3 introduces the extracted visual features, including 2D Zernike moments, 2D Fourier descriptor and 2D Krawtchouk moments. Multi-modal graph learning based feature fusion is presented in detail in Section 4. Experiments conducted on ETH and NTU 3D datasets, the experimental results and analysis are provided in Section 5, followed by conclusion and future work in Section 6.

2. Related work

View representation: The Zernike moments and Fourier descriptor are the most widely used features to represent views of 3D objects [7,10,17]. Daras et al. proposed a unified framework for 3D shape retrieval, which supports multimodal queries by introducing a view-based approach [10]. In the method, multiple views of a 3D object are generated by taking views from uniformly distributed viewpoints and a set of 2D rotation-invariant shape descriptors, including the Polar-Fourier Transform, Zernike Moments and Krawtchouk Moments, are produced to construct a Compact Multi-View Descriptor. LightField Descriptor (LFD) is used to capture representative views, which is defined as features of 10 images fendered from vertices of dodecahedron over a hemisphere [7]. To improve the robustness, a set of light fields is applied in [7]. Elevation Descriptor (ED) [15] is proposed for 3D model retrieval. Six elevations are obtained to describe

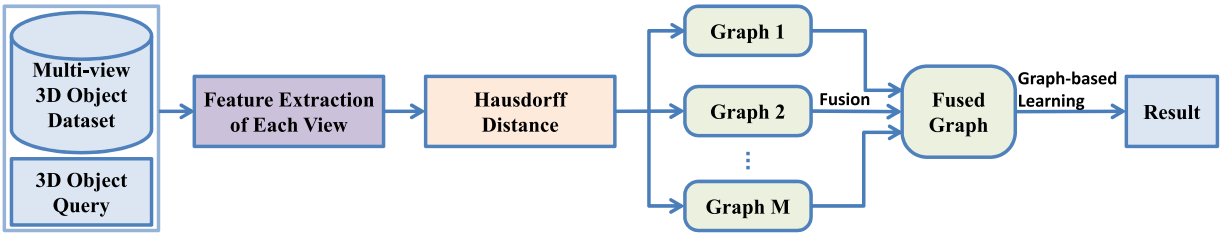

Fig. 1. The framework of the proposed feature fusion method based on multi-modal graph learning for view-based 3D object retrieval.

Download English Version:

https://daneshyari.com/en/article/562493

Download Persian Version:

https://daneshyari.com/article/562493

Daneshyari.com