Graphical Models 85 (2016) 22–29

Contents lists available at ScienceDirect

Graphical Models

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

Upright orientation of 3D shapes with Convolutional Networks

Zishun Liu*, Juyong Zhang, Ligang Liu

University of Science and Technology of China, Hefei, Anhui, China

ARTICLE INFO

Article history:
Received 30 November 2015
Revised 1 March 2016
Accepted 6 March 2016
Available online 14 March 2016

Keywords:
Upright orientation
Data-driven shape analysis
Voxelization
Convolutional Networks

ABSTRACT

Posing objects in their upright orientations is the very first step of 3D shape analysis. However, 3D models in existing repositories may be far from their right orientations due to various reasons. In this paper, we present a data-driven method for 3D object upright orientation estimation using 3D Convolutional Networks (ConvNets), and the method is designed in the style of *divide-and-conquer* due to the *interference effect*. Thanks to the public big 3D datasets and the feature learning ability of ConvNets, our method can handle not only man-made objects but also natural ones. Besides, without any regularity assumptions, our method can deal with asymmetric and several other failure cases of existing approaches. Furthermore, a distance based clustering technique is proposed to reduce the memory cost and a test-time augmentation procedure is used to improve the accuracy. Its efficiency and effectiveness are demonstrated in the experimental results.

1. Introduction

Most objects are usually posed in their upright orientations, which makes them easily recognizable. Also, it is the very first step to pose the given 3D shapes in their upright orientations (Fig. 1) in many graphics and robotics tasks, such as matching [2], retrieval [13,28], shape analysis [34] and placement planning [17]. Moreover, it can be used to generate recognizable object thumbnails, helping the management of 3D shape repositories. Due to various reasons such as modeling platforms or scanning systems, many models in existing databases are not in their upright orientation. Therefore, a number of approaches have been proposed to handle this problem. However, these methods are usually limited to shapes with some regularity and take several seconds to process each shape. Thus more efficient and effective methods are needed.

In this paper, we present a learning based method to predict the upright orientation using 3D Convolutional Networks (ConvNets). Given voxel representations of 3D shapes and corresponding orientation vectors, this prediction task can be formulated as a regression problem. Leveraging the learning ability of deep neural networks, general categories of 3D shapes can be handled without making any assumptions such as symmetry or parallelism. Besides mesh models, the proposed method can deal with shapes represented in other types that can be voxelized, such as implicit surfaces and point clouds, without surface reconstruction [7].

Compared with the ConvNets based approach, existing methods are limited by their predefined rules. For example, the method proposed by Fu et al. [8] is based on the observation that man-made object should have a supporting base on which it can be steadily positioned. Nevertheless, this observation is not applicable to all shapes, especially natural ones. Thus learning based methods are appreciated to deal with general objects. Although the idea of data-driven is adopted in Fu et al. [8], the learning procedure is based on the hand-crafted features such as

* Corresponding author.
E-mail address: zishun@mail.ustc.edu.cn (Z. Liu).

http://dx.doi.org/10.1016/j.gmod.2016.03.001
1524-0703/

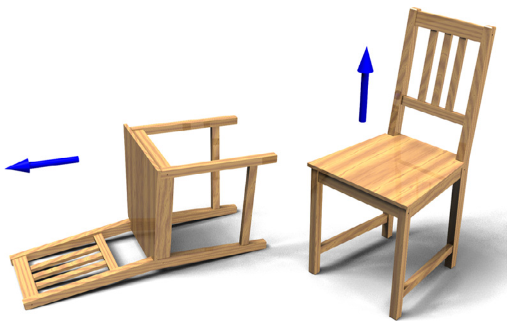

Fig. 1. Upright orientation estimation.

stability, visibility and parallelism, which fall into the field of feature engineering. In one word, it is hard to define a universal rule to upright general 3D shapes effectively. By contrast, neural networks work in the style of end-to-end learning. High-level knowledges can be captured from raw data, without relying on object's regularity such as explicit symmetry.

However, a single ConvNet does not work well for all types of shapes. The key challenge is that each shape category exhibits particular characteristic on the upright orientation, for example, cars tend to be horizontal while bicycles are likely to be vertical. This is referred to as *interference effect* [14] which will lead to poor generalization. In other words, different strategies should be taken to handle diverse categories. Thus a *divide-and-conquer* scheme is used in our system. Each shape is first classified by a network and then fed into one of the orientation regression networks that are trained on each of the categories. Furthermore, a distance based clustering method is proposed to reduce the number of networks and a novel test-time augmentation procedure is used to improve the accuracy.

The efficiency and effectiveness of this approach are demonstrated by extensive experiments. Our system achieved the accuracy of more than 90% on the test data and showed the generalization capability of inferring upright orientations for shapes not belonging to the training categories. Also experimental results showed that our system is able to handle several cases that other methods fail. Moreover, estimation for each shape took no more than 0.15 s on average, which is much faster than existing approaches, thus applicable to robotics tasks in which immediate feedback is required.

The main contributions of our approach are summarized in the following.

- General objects can be handled by this approach thanks to the learning ability of ConvNets, including asymmetric shapes.
- The proposed method is at least 30 times faster than existing methods.

The remainder of this paper is structured as follows. Section 2 briefly reviews several related works. In Section 3 our network system is specified. The experimental results and comparisons with related works are demonstrated in Section 4. Finally, Section 5 presents our conclusions and directions of future work to improve our method.

2. Related work

Orientation of images. Images may differ from their correct orientations by 0°, 90°, 180°, or 270° [3,23,24,32]. Therefore, the image orientation detection problem can be formulated as a four-class classification problem. Most of the existing approaches extract high dimensional feature vector in each possible orientation and then train support vector machines (SVM) [23,32] or other classifiers [3] on feature vectors to detect correct orientation. However, it is difficult to reduce the two-dimensional orientation space to a few candidates for general 3D objects. Thus we formulate the upright orientation estimation of 3D models as a regression problem.

Upright orientation of 3D models. In computer graphics, several methods have been proposed to estimate upright orientation or align the given models. One commonly used method is the principal component analysis (PCA) [19] which is inaccurate and not robust for many models, especially asymmetric ones. In Fu et al. [8] and Lin and Tai [22], upright orientation is estimated using supporting base candidates on which a 3D model can stand upright. These methods work well for most of the man-made models while not applicable to natural objects whose supporting bases are not well defined. Another type of method is based on the observation that the coordinate matrix of the 3D object with upright orientation should have reduced rank. Inspired by [37], Jin et al. [18] present an algorithm in which a 3D shape is aligned with axes by iterative rectification of axis-aligned projections as low-rank matrices independently. In Wang et al. [31], a method is proposed by minimizing the tensor rank of the 3D shape's voxel representation. Both methods can handle shapes that have some kinds of symmetries. We can see that none of the above methods is able to deal with general objects.

Viewpoint selection. Representative viewpoint provides the most informative and intuitive view of a 3D shape, which benefits many geometry processing applications like shape retrieval. Most approaches select representative viewpoints using geometric information of the 3D models, such as number of visible polygons [25] and silhouette contours [1]. Some works are based on information theory, such as viewpoint entropy [29], multi-scale entropy [30], and viewpoint mutual information [6]. It will be much easier to select the representative views for 3D models if they are posed at the upright orientation by our method.

3D shape matching, retrieval and registration. 3D shape retrieval [13,28] and matching [2] techniques attempt to find the similar shapes from databases with queries. 3D shape registration techniques [36] make efforts to find corresponding parts of multiple models. These methods are trying to design a robust and efficient method for measuring the similarity between two shapes or parts over the space of all transformations [19]. To address this issue, most techniques pre-align the models into a common coordinate frame, typically using PCA alignment.