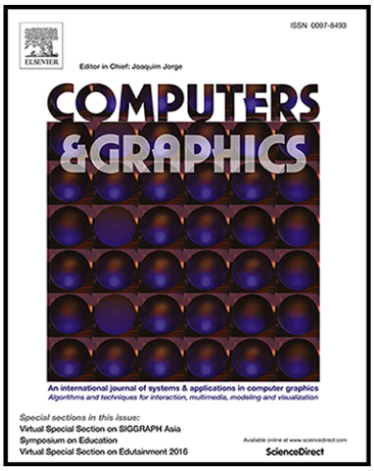
Accepted Manuscript



A Multi-view Recurrent Neural Network for 3D Mesh Segmentation

Truc Le, Giang Bui, Ye Duan

PII: S0097-8493(17)30060-2
DOI: 10.1016/j.cag.2017.05.011
Reference: CAG 2789

To appear in: *Computers & Graphics*

Received date: 30 March 2017
Revised date: 17 May 2017
Accepted date: 25 May 2017

Please cite this article as: Truc Le, Giang Bui, Ye Duan, A Multi-view Recurrent Neural Network for 3D Mesh Segmentation, *Computers & Graphics* (2017), doi: 10.1016/j.cag.2017.05.011

A Multi-view Recurrent Neural Network for 3D Mesh Segmentation

Truc Le, Giang Bui and Ye Duan

Department of Computer Science, University of Missouri - Columbia

Abstract

This paper introduces a multi-view recurrent neural netowrk (MV-RNN) approach for 3D mesh segmentation. Our architecture combines the convolutional neural networks (CNN) and a two-layer long short term memory (LSTM) to yield coherent segmentation of 3D shapes. The imaged-based CNN are useful for effectively generating the edge probability feature map while the LSTM correlates these edge maps across different views and output a well-defined per-view edge image. Evaluations on the Princeton Segmentation Benchmark dataset show that our framework significantly outperforms other state-of-the-art methods.

Keywords: mesh segmentation, multi-view, 3D deep learning, CNN, RNN, LSTM

1. Introduction

Mesh segmentation is a classical, yet challenging problem in computer graphics for many decades. Unfortunately, the segmentation problem is ill-posed and there is no general objective measurement that can universally be applied in any case. Judging the quality of a segmentation largely depends on application. For instance, in a LiDAR scan of urban environment, a desired segmentation should distinguish between different instances of buildings, people, cars, trees, ground, etc. However, in a part-based annotation of a 3D model (e.g. human), the requirement is usually to segment head, torso, left/right arms, left/right legs and sometimes more details such as thumb, index finger, and so on depending on specific task. Consequently, in the scope of this paper, we aim to tackle the mesh segmentation as a data driven approach. Given a training dataset of input mesh and the corresponding desired segmentation, we design a deep learning framework to learn the pattern of the segmentation given by the training dataset so that it can segment an unseen mesh. As a result, we make no geometric or topological assumptions about the shape, nor exploit any hand-crafted descriptors.

In this paper, we propose a multi-view recurrent neural network (MV-RNN) deep learning framework to segment 3D model which significantly outperforms prior methods on the Princeton Segmentation Benchmark dataset [1]. It is worth mentioning that our goal is to partition the 3D model and not to do the semantic segmentation. In semantic segmentation, the two wings of an airplane are assigned a single label *wing*. On the other hand, in mesh segmentation, the two wings belong to two different regions and do not have semantic label. In general, semantic segmentation provides better understanding of a 3D model. However, mesh segmentation still has its merits such as guiding mesh processing algorithms including skeleton extraction [2, 3], modeling [4], morphing [5], shape-based retrieval [6] and texture mapping [7]. Moreover, in contrast to semantic segmentation which requires a fixed set of semantic labels, many mesh segmentation algorithms could be generalized to unseen object categories. As a result, instead of identifying surface area of the 3D model within a segment, we predict its boundary (or edge). The benefits of doing so are twofold. First, it is usually more expensive to obtain dense surface annotations than boundary annotations from humans. Second, we only have two semantic labels, i.e. boundary versus non-boundary, which is simpler for the framework to learn than using hundreds of semantic labels (e.g. hand, torso, leg, head, etc.). In fact, detecting 3D edges could be useful for other tasks such as suggestive contours [8, 9] and ridge-valley detection [10].

Our approach belongs to the multi-view paradigm which has been shown success recently for many visual recognition tasks such as classification and segmentation [11, 12, 13, 14]. Typically, in the multi-view segmentation, a 3D model is rendered with multiple views to generate multi-view images, each of which is fed-forward to a (shared weights) convolutional neural network to obtain densely labeled images before being mapped back to 3D. In general, a multi-view approach for segmentation must overcome several technical obstacles. Firstly, there must be enough views to minimize occlusions and cover the shape surface. This can be achieved by generating a large number of views equally distributed around the object. Secondly, shape parts can be visible from more than one view, thus, the proposed method should effectively correlate information from multiple views. The main drawback of the existing multi-view approaches such as the multi-view convolution neural network (MV-CNN) [11, 12] is that different views may not be correlated and hence a 3D area may correspond to totally different outcomes from different views. Let us take an example of a standing person rotating counter-clockwise (Fig. 1). When the view is front facing, the boundary between the torso and the right arm is a real boundary. At certain time, the right arm starts to be occluded. Then the boundary between the torso and the right arm is no longer real boundary, but the MV-CNN cannot distinguish them because it does not correlate the result over