3D Keypoint detection with Deep Neural Networks

Elías J. Puma

Escuela Profesional de Ciencia de la Computación,
Universidad Nacional de San Agustín,
Arequipa, PE
Email: eliasj.puma@gmail.com

3D keypoint detection plays a fundamental role in the Computer Vision field, detection of these salient points in the local surfaces of a 3D object is important in order to perform certain tasks such as registration, retrieval and simplification. There has been a lot of research in the field of 3D keypoint detection, most of them take a geometrical approach which have a good performance but lack flexibility to adapt to changes such as noise and high curvature points that are not keypoints to human preference. A good approach seems to be machine learning methods that can be trained with human annotated training data. In this paper a new method is proposed using deep neural network with sparse autoencoder as the regression model due to their great ability for feature processing. The analysis shows this method would outperform other methods that are widely used.

Keywords: Keypoint detection; Deep Neural Networks; 3D Model; Sparse Autoencoders

1. INTRODUCTION

Several computer-dependent areas are benefited of the applications that 3D Models have in them. The growth of 3D data has increased in the last years with the availability of low-cost 3D capture devices [1]. The ability to analyse, proccess and select relevant information from them is an active research area.

1.1. Keypoint detection

There is not a definition of what exactly a keypoint is, but a common assumption relates the level of protusion of outstanding local structures with the measure of interest of such local structure [1]. We can say then, that a flat or almost flat surfaces aren't keypoints, as neither the edges of an object are. Its main caracteristic is its invariance to transformations in the object itself.

1.2. Approaches

3D interest point detection is a difficult task for several reasons [3, 1]. First, there are not any definitions for what a interest point is, most of the approaches consider the high level of protusion in a local area as a keypoint characteristic. So, in planar sections of an area vertices have a low interest level and in local areas with diferent structures the interest level will be the opposite. Second, vertex density is different for every 3D model which makes harder the task of selecting a local area. Third, information obtained from a 3D model are only vertex positions and connectivity between them which means the interest level will depend only from the information we can retrieve from different calculations. These are not the only reasons but are sufficient for explaining why this method is prepared to handle these difficulties.

The common approach to 3D keypoint detection has been to use geometric properties of the models, although in recent years researchers also have developed machine learning techniques that try to outperform the former one by avoiding the problems of: Different tasks in different areas of the model [2], false positives obtained from noise or local variation and keypoint detection valuable according to human preference.

The rest of this paper is organized in the following way: Section 2 introduces previous work done in the area, Section 3 presents the idea this method is based on. Future results will be presented in Section 4 and future conclusions in Section 5.

2. RELATED WORK

In recent years researchers have proposed several techniques for 3D keypoint detection. Most of them are based on geometric methods, for instance Sipiran and Bustos extended the Harris operator for 3D meshes [1]. Lin, Zhu, Zhang and Liu proposed a geometric technique [5] based in the tangencial planes traced for each vertex and other transformations in the mesh some of them can be also found in [2].

The lack of flexibility of geometric methods led researchers to find new methods to detect 3D keypoints.

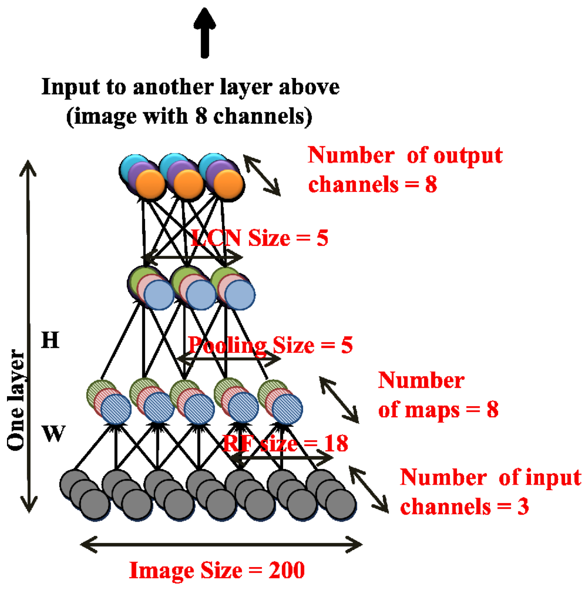

FIGURE 1. Large scale unsupervised learning architecture [4]

3. PROPOSAL

This work is inspired by the work of [2] by the use of sparse autoencoders as the regression model in order to learn features from local and global information generated from a human-annotated keypoint database. Also this proposal is influenced by [4], so for us to achieve the 3D keypoint detection we train a 3-layered locally connected sparse autoencoder similarly to their technique, such technique's results revealed to be a inexpensive way to develop high-level features from unlabeled data, from that study this work presents an adapted architecture for 3D meshes. Taking a different approach than the other techniques mentioned above, resized and simplified segments of the 3D mesh will be used as the input for our Deep Neural Network, this will enable our DNN to learn when one of these segments has inside a keypoint.

3.1. Architecture

This technique can be seen as a set of sparse deep autoencoders that similarly to [4] has two fields in it: local receptive fields, pooling normalization (the architecture taken as a base can be seen on the Figure 1). Local receptive fields scale the autoencoder to big inputs, connecting the autoencoder's features to a small region of the next lower layer. These sublayers are know as filtering and pooling.

Originally the neurons in the first sublayer were connected to pixels in all input channels [4], but in order to adapt this architecture it is proposed to use the 3D vertices and their connectivity information as the input channels and by so adding more receptive fields.

3.2. Training

As mentioned before the first layer input...

To train the Deep Neural Network what is to be done at first is to train each Sparse Autoencoder and a final logistic regression layer, then following the schema from [2] stack the four layers together and backpropagate the whole DNN to fine tune it.

The goal of this approach is to reduce the proccessing that is performed, instead of evaluating each vertex in the DNN which is expensive, we can perform the neccesary calculations just for some samples of the 3D object and discard if those samples don't contain any keypoints, in the case we find the presence of keypoints we will perform further calculations to choose the sample keypoint.

4. RESULTS

5. CONCLUSIONS

ACKNOWLEDGEMENTS

This research was undertaken as part of ...

REFERENCES

[1] I. Sipiran, B. Bustos (2011) Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes Santiago, Chile: Springer, 2011.

[2] X. Lin, C. Zhu, Q. Zhang y Y. Liu (2016) 3D Keypoint Detection Based on Deep Neural Network with Sparse Autoencoder Chengdu, China: UESTC.

[3] L. Teran, P. Mordohai (2013) 3D Interest Point Detection via Discriminative Learning Stevens Institute of Technology, EUA, 2013.

[4] Quoc V. Le, Marc' Aurelio Ranzato, Andrew Y. Ng (2012) Building High-level Features Using Large Scale Unsupervised Learning Proceedings of the 29t̂h International Conference on Machine Learning, UK, 2012.

[5] X. Lin, C. Zhu, Q. Zhang y Y. Liu (2016) 3D Interest Point Detection Based on Geometric Measures and Sparse Refinement Chengdu, China: UESTC, 2016.