GFPNet. Neural network based volumetric object reconstruction using generic fitted primitives

Alexandru Razvant, Tiberiu Teodor Cocias, Sorin Grigorescu

The authors are with Elektrobit Automotive and the Robotics, Vision and Control Laboratory (ROVIS Lab) at the Department of Automation and Information Technology, Transilvania University of Brasov, 500036 Romania. E-mail: (see http://rovislab.com//tiberiu_cocias.html).

1 INTRODUCTION

3D volume reconstruction from partial point clouds remains still one of the fundamental problem in perception systems. Volumetric reconstruction is usually achieved by applying formal methods that fits a volume onto the partially viewed object. The goal is to obtain a volume of the object that is closest to its true form. Data driven approaches for volumetric reconstruction are the most encountered in the literature. In [2] the authors propose a reconstruction apparatus that first detects objects and estimates their pose coarsely in the stereo images using state-of-the-art 3D object detection method. Following, an energy minimization method then aligns shape and pose concurrently with the stereo reconstruction of the object. The shape is retrieved from a shape database. The process of retrieving the optimal shape is slow and makes the approach nearly un-usable in real-time systems. Artificial intelligence may play a key role in improving the performance of such algorithms. Learning techniques such as End2End or Deep Reinforcement Learning (DRL) can be used to create a pattern of how the full representation of a 2.5D object may be reconstructed. In [1] the authors propose a weakly-supervised learning-based approach to 3D shape completion which neither requires slow optimization nor direct supervision. The approach also learn a shape prior on synthetic data in order to amortize, ie, learn, maximum likelihood fitting using deep neural networks resulting in efficient shape completion without sacrificing accuracy.

In this paper we propose a 2 step volumetric object reconstruction framework that uses a neural network in order to improve the appearance of a generic volume fitted on the partially viewed object. First we fit a generic primitive (GP) onto the 2.5D perceived object to obtain an initial volume. Second, the generic primitive is modelled using a deep neural network (DNN) in order to capture the particularities of the perceived object. The modelling process is performed for each point of the primitive. The neural network behaves like a solver that tries to optimize a multi-objective fitting function to ensure that the geometry, consistency and smoothness of the priomitive surface is as close as possible to the real object. Tackling 3D volume reconstruction of the objects from the KITTI dataset [3], we demonstrate that the proposed reconstruction framework is able to compete with a fully supervised baseline and a state-of-the-art data-driven approach.

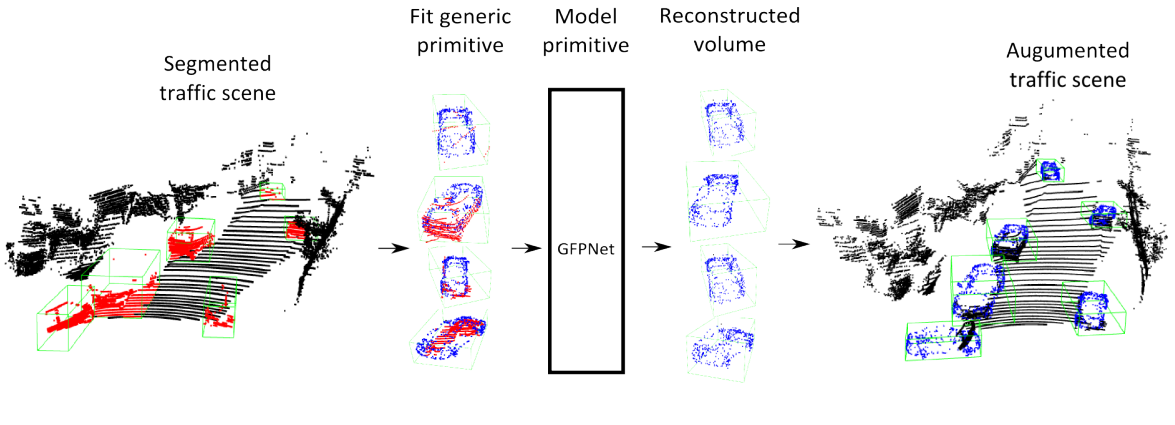

Figure 1: **Overview of the volumetric reconstruction framework**. The leftmost image describes a LIDAR scan of a traffic scene from the KITTI dataset. The first column with images presents the result of the fitting process between a car primitive and the segmented car point cloud. The second column with images present the modelled primitive while the rightmost image depicts the overlay of the reconstructed car volumes onto the original scene.

2 METHOD

The first step in the proposed framework is to segment the objects from the scene that are placed on flat surfaces (e.g. on the road). As a result of segmentation, different 3D object clusters C_k, called also point density models

(PDMs), are obtained. On each detected object we apply a neural network based 3D point cloud classification approach to determine the object's class [4]. Further, a generic primitive (GP) of that class is fitted onto the PDM using the iterative closes point (ICP) approach. In the last step we use a deep neural network to minimize a multi-objective fitting function that deforms the GFP points such that it captures the particular shape of the real object in the regions where the information was available. The overview of the GFP approach is shown in Fig. 1.

2.1 Generic primitive definition

A generic primitive S is considered to be a volume constructed in such a manner that it resembles many similar objects from the same class. Each class has his own generic primitive. E.g. different types of bottles can be approximated by a common generic primitive of a bottle. The shape of a generic primitive is obtain using *Generalized Procrustes Analysis* [5] by averaging the shapes of the objects from a class.

2.2 GFPNet overview

Once the primitive is aligned with the perceived object cloud, the proposed approach aims at sequentially model each primitive point p_i^s using the GFPNet neural network. Through this step we obtain predictions of the positions of p_i^s and 5 of his primitive neighbours. The architecture of the GFPNet deep neural network is presented in Fig. 2. The input tensor fed to the network has the shape $Nx6$, where N is the number of 3D points considered and 6 is the Cartesian position x,y,z and normal n_x,n_y,n_z of each 3D point. N has a fixed size of 30 and it contains: (a)current primitive point p_i^s, (b)5 primitive neighbours points that are around p_i^s (determined using the Delaunay triangulation approach) and (c)24 neighbours points from the perceived object cloud that are around p_i^s (determined using k nearest neighbour (KNN) radius search). The output tensor has a shape of $6x3$, which depicts 6 primitive points and their Cartesian coordinates. The first point is p_i^s, and the rest are it's 5 neighbours.

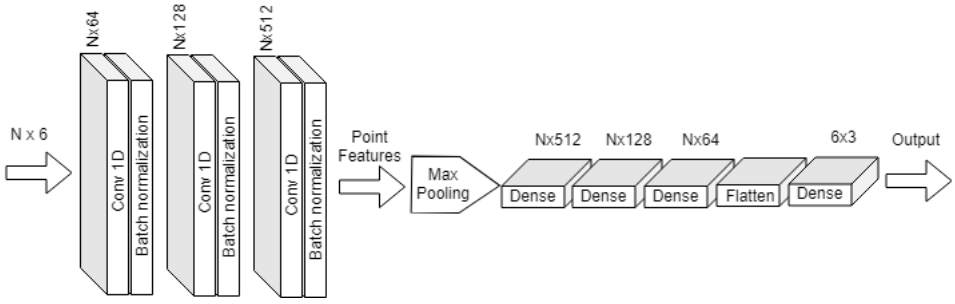

Figure 2: **GFPNet architecture**.

3 RESULTS

We tested the GFPNet on the KITTI dataset [3] in order to reconstruct the volumes of trucks, vehicles and pedestrians. For training the network we have used 621 different objects (356 vehicles, 135 pedestrians and 130 trucks) out of which we have extracted 187346 training samples. We trained the GFPNet over 1000 epochs with a batch size of 512 using ADAM optimizer with a learning rate of 0.001 and a decay of 0.9.

For labelling the data a semi-automatic approach was used. Based on our previous work [6], we first apply a formal approach to model each primitive point. Second, using a visual inspection tool we manually adjust the previous step results in order to refine the output. The criteria for modifying the 3D position of a primitive point are: (1)the euclidean distance d between p_i^s and the neighbour object cluster points p_j^c in a given radius around p_i^s and (2)the smoothness λ factor must be minimum. The λ factor is determined using the first and second derivative of the position of the primitive point relative to it's neighbour primitive points. E.g. λ is 0 when the surface is flat or it is smooth.

4 CONCLUSION

In this paper, a deep neural network shape modelling framework has been presented in the context of traffic scene augmentation. It's main goal is to deliver precise 3D volumetric estimation of the objects and obstacles present in the traffic scene. The main novelty of the algorithm lies in the point wise surface modelling of a generic volume of

an object using a deep neural network. Future work aims at gradually increasing the surface that is being modelled till in the end the global volume is considered.

References

[1] D. Stutz and A. Geiger, "Learning 3d shape completion from laser scan data with weak supervision," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2018.

[2] F. Engelmann, J. Stückler, and B. Leibe, "Joint object pose estimation and shape reconstruction in urban street scenes using 3d shape priors," vol. 9796, 09 2016, pp. 219–230.

[3] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[4] R. Qi Charles, H. Su, M. Kaichun, and L. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," 07 2017, pp. 77–85.

[5] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis, with Applications in R. Second Edition.* Chichester: John Wiley and Sons, 2016.

[6] T. T. Cocias, F. Moldoveanu, and S. M. Grigorescu, "Generic fitted shapes (GFS): volumetric object segmentation in service robotics," *Robotics and Autonomous Systems*, vol. 61, no. 9, pp. 960–972, 2013. [Online]. Available: https://doi.org/10.1016/j.robot.2013.04.020

[7] L. A. Marina, B. Trasnea, T. T. Cocias, A. Vasilcoi, F. Moldoveanu, and S. M. Grigorescu, "Deep grid net (DGN): A deep learning system for real-time driving context understanding," in *3rd IEEE International Conference on Robotic Computing, IRC 2019, Naples, Italy, February 25-27, 2019*, 2019, pp. 399–402. [Online]. Available: https://doi.org/10.1109/IRC.2019.00073