Automatic semantic segmentation of 3D urban scenes

Federico Tombari
DEIS-ARCES
University of Bologna, Italy
federico.tombari@unibo.it

Luigi Di Stefano
DEIS-ARCES
University of Bologna, Italy
luigi.distefano@unibo.it

1. Introduction

In this short paper we briefly present the results concerning automatic semantic segmentation of a public 3D dataset. This dataset concerns urban scenes, was acquired with a lidar sensor and was proposed for the *3DIMPVT 2011 Urban Data Challenge*. The problem that is tackled here is automatic segmentation of urban scenes acquired with a 3D sensor into major categories of interest, such as buildings, vehicles, vegetation [3]. The algorithm employed to achieve the results is described in detail in [4].

First, in Section 2, we provide an overview of the segmentation algorithm. Then, in Section 3 we describe in detail the dataset used in the experimental evaluation. Finally in Section 4 we discuss the qualitative results yielded by our approach.

2. Algorithm

The algorithm used in our experiments is proposed in [4]. This algorithm is designed to perform segmentation of 3D data into rigid/deformable object instances as well as object categories. It assumes a *learning* stage, performed *off-line*, aimed at training a multi-class classifier by means of labeled training data. As the first step, 3D feature detection and description is carried out on each 3D model to obtain descriptions of salient 3D keypoints. Successively, extracted descriptors are fed to a multi-class classifier as training data.

During the *on-line* stage, an identical 3D feature detection and description process is run on the 3D scene to be segmented. Then, feature points are classified based on their descriptors. Next, labeled features are fed to a grouping stage based on a Markov Random Field formulation which attains the final segmentation starting from the initial labeling provided by the local classifier. More specifically, during this grouping stage, each labeled feature represents a node over an undirected graph. Minimizing a global energy through a MRF formulation allows for enforcing local consistency between classified labels, under the reasonable assumption that neighboring features tend to share the same category.

The algorithm can work with linear as well as non-linear classifier. Training time can be very small since suitable classifiers (e.g. Nearest-Neighbor) can be deployed while the grouping stage does not need any training. Results shown in [4] show that this approach obtains high segmentation accuracy with several popular classifiers such as Support Vector Machines (SVM) [1], Boosting [2] and NN [5].

3. Dataset

The dataset used to obtain the results shown in this report is one of the two datasets made available for the 3D Urban Data Challenge, and is referred to as *New York City* (NYC). After a first analysis of the dataset, we dived it into three main categories: *facades*, *vegetation*, *vehicles*. The original dataset is subdivided into files having a huge difference in the part of the city being represented (e.g. from a huge street with several buildings down to a small fragment of a single building). This also accounts for a huge difference in the file sizes. We selected a subset of the dataset composed of 11 scenes, by cutting them off of the biggest files, making sure that each of the 11 scenes included - at least in small parts - all the 3 data categories. Out of these 11 scenes, the first one is used for validation, while the remaining ones are used as testing set. As for the training set, we selected from the remaining data (i.e. those not included in the testing set) 5 parts per class (for a total of 15 parts) that were used to train our classifier.

4. Experimental results

We have included the experimental results concerning the evaluated algorithm over the NYC dataset in the supplementary material of this submission [1]. All the results shown are qualitative, due to the lack of ground-truth. In the results, we first show the full training dataset, with 5 models for category. due to the small size of the deployed

[1] Available at `vision.deis.unibo.it/fede/3Dsegm.html`

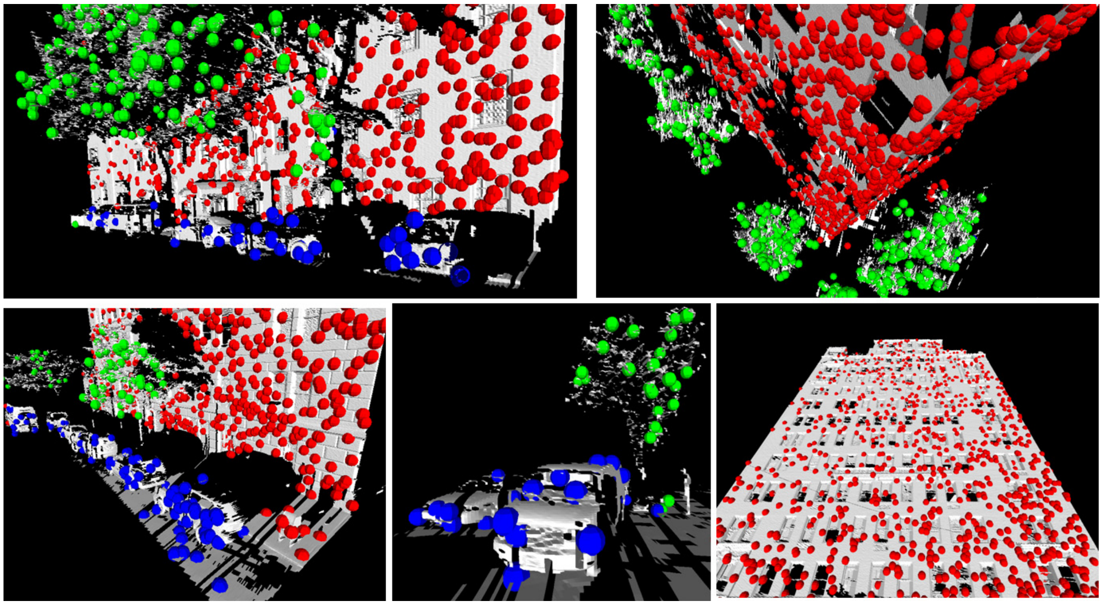

Figure 1. Some details of the final segmentation taken throughout the dataset.

training set and to the small number of categories, certain parts (such as particular building entrances) and objects (eg. lamps, posts, ..) were not taken into account in the learning stage, although they are particularly recurring in the test set. Hence, these are the areas of the scenes were the segmentation yielded by the algorithm is less reliable. Nevertheless, this can be easily taken into account if more training data are available.

Successively, we show the scene used for validation and the results of three classifiers (NN, SVM, Boost) over this scene. Each category is represented by a different color (red: facades, blue: vehicles, green: vegetation). Results are shown both before (left) and after (right) the grouping stage applied by the algorithm. In addition, we show the result obtained by the proposed algorithm on the test set using, respectively, a SVM classifier, a NN classifier and a Boost classifier. Each figure is composed of 10 sub-figures, each of which depicts, from left to right, the scene currently tested, the partial results yielded by the algorithm after classification and before grouping, and the final results obtained after the grouping stage. Finally, to better highlight the precision in the segmentation yielded by the algorithm, we show some details of the final segmentation yielded by the algorithm throughout the dataset. A subset of snapshots of the final segmentation is also shown in Figure 1.

The proposed results demonstrate that the yielded segmentation is overall accurate, and can robustly deal with the intra-class variations present throughout the test dataset. Also, the grouping stage seems effective in improving the results of the classifier. Finally, the proposed algorithm obtains accurate (and almost equivalent) results with all deployed classifiers, though SVM seems the classifier yielding, overall, the best qualitative results.

5. Final remarks

The dataset used for the experimental evaluation does not come with texture. Hence, only the shape cue has been exploited to perform semantic segmentation. Nevertheless, the adopted algorithm [4] can exploit jointly texture and shape so as to achieve a more accurate segmentation in case of textured 3D datasets.

6. Acknowledgments

The authors would like to thank Simone Giardino, who helped creating the dataset and running the experiments.

References

[1] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[2] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Proc. EuroCOLT*, pages 23–37, 1995.

[3] D. Munoz, J. A. Bagnell, N. Vandapel, and M. Hebert. Contextual classification with functional max-margin markov networks. In *Proc. CVPR*, 2009.

[4] F. Tombari and L. Di Stefano. 3d data segmentation by local classification and markov random fields. In *3DIMPVT 11, accepted*, 2011.

[5] R. Triebel, R. Schmidt, O. M. Mozos, and W. Burgard. Instance-based amn classification for improved object recognition in 2d and 3d laser range data. In *Proc. Int. J. Conf. on Art. Intelligence*, 2007.