3D All The Way: Semantic Segmentation of Urban Scenes From Start to End in 3D

Anđelo Martinović♠[1], Jan Knopp♠[1], Hayko Riemenschneider[2] and Luc Van Gool[1,2]
[1]PSI-VISICS, KU Leuven, Belgium. [2]Computer Vision Lab, ETH Zurich Switzerland.

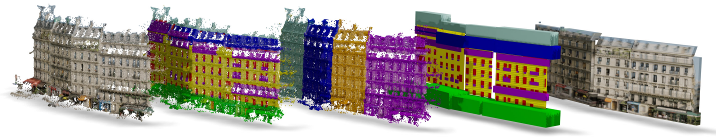

Figure 1: **Our approach.** In contrast to the majority of facade labeling methods, our approach operates completely in 3D space. (a) image-based SfM 3D point cloud (b) initial point cloud classification (c) facade splitting (d) structure modeling through architectural principles and (e) projected original images onto estimated 3D model. The advantages of pure-3D range from tremendous speed-up to complementarity with 2D classifiers.

We propose a new approach for semantic segmentation of 3D city models. Starting from an SfM reconstruction of a street scene, we perform classification and facade splitting purely in 3D, obviating the need for slow image-based semantic segmentation methods. Our properly trained pure-3D approach produces high quality labelings, with significant speed benefits (20x faster) allowing us to analyze entire streets in a matter of minutes. Additionally, if speed is not of the essence, the 3D labeling can be combined with the results of a state-of-the-art 2D classifier, complementing the performance. Further, we propose a novel facade separation based on semantic nuances between facades. Finally, inspired by the use of architectural principles for 2D facade labeling, we propose new 3D-specific principles and an efficient integer quadratic programming optimization.

	Low-res PCL		High-res PCL	
Method	Accuracy	Timing	Accuracy	Timing
Bottom-up semantic segmentation				
Select 2D [2]	42.32	15min	39.92	23min
Full 2D [1]	55.72	302min	54.30	305min
Our 3D	52.09	15min	56.39	76min
Our 3D+2D [1]	60.43	397min	61.39	470min
Weak architectural rules on top of Ours+Layer 1				
2D Rules [1]	58.81	802min	58.40	885min
Our 3D Rules	60.83	8min	59.89	10min

Table 1: **Semantic segmentation of point clouds:** Pascal IoU accuracy and timing on the RueMonge2014 dataset. Our approach provides a very fast alternative to previous methods, while providing top results.

Our goal is to estimate a semantically segmented 3D scene starting from images of an urban environment as the input. As a first step, we obtain a set of semi-dense 3D points from standard SfM/MVS algorithms. Next, we classify each point P_i in the point cloud into one semantic class L_i (window, wall, balcony, door, roof, sky, shop), using a Random Forest classifier trained on light-weight 3D features. Afterwards, we separate individual facades by detecting nuances in semantic scene understanding (Fig. 1, middle). Finally, we propose architectural rules that preserve principles such as the alignment or co-occurrence of facade elements. These rules have two effects: they improve our–already very good–results, and directly return the high-level 3D facade structure (Fig. 2, bottom).

[1] Anđelo Martinović, Markus Mathias, Julien Weissenberg, and Luc Van Gool. A three-layered approach to facade parsing. In *ECCV*, 2012.

[2] Hayko Riemenschneider, Andras Bodis-Szomoru, Julien Weissenberg, and Luc Van Gool. Learning Where To Classify In Multi-View Semantic Segmentation. In *ECCV*, 2014.

♠ Indicates equal contribution. This work was supported by the KU Leuven Research Fund, iMinds, FWO-project G.0861.12, and European Research Council (ERC) projects COGNIMUND (#240530) and VarCity (#273940). The code is available at https://bitbucket.org/amartino/facade3d. This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

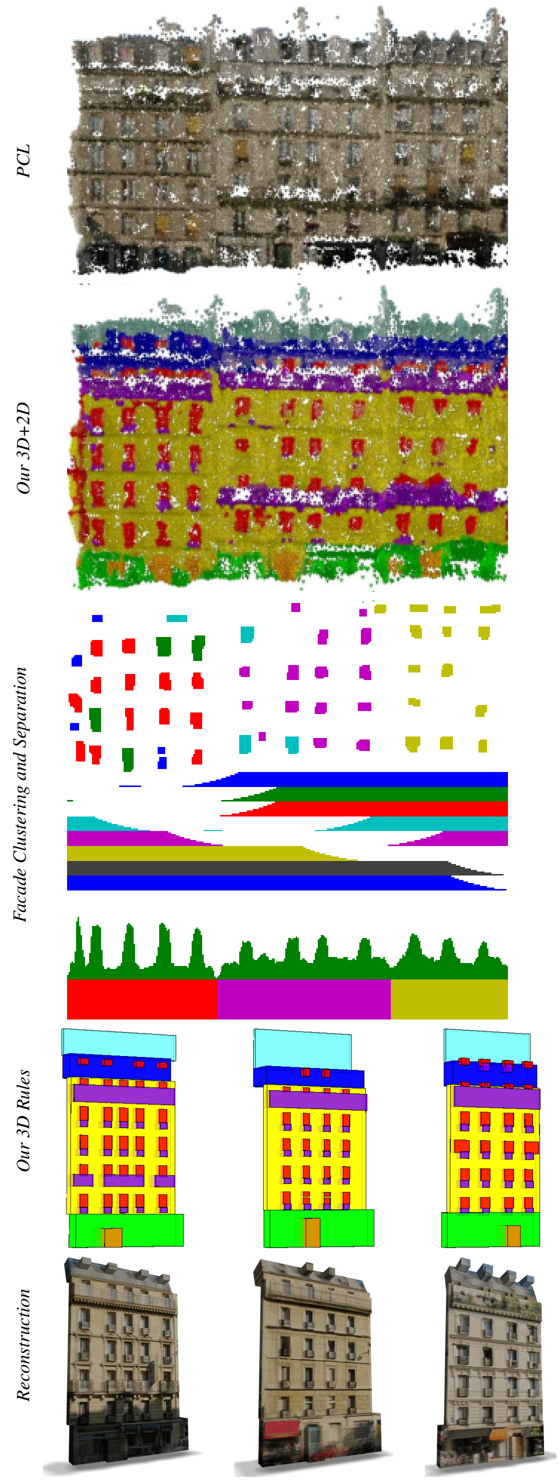

Figure 2: Qualitative results of reconstructed facades using our method. From top to bottom: initial colored point cloud (Low-res), semantic classification, facade element clustering and separation, estimated boxes using weak 3D rules; and automatically generated 3D model of the facade with projected ortho images.