Shape Grammars and Procedural Modeling towards Large Scale 3D Modeling and Reconstruction

Loic Simon[1]
http://www.mas.ecp.fr/vision/Personnel/simon
Olivier Teboul[1,3]
http://www.mas.ecp.fr/vision/Personnel/teboul
Panagiotis Koutsourakis[1,4]
panagiotis.koutsourakis@ecp.fr
Iasonas Kokkinos[1,2]
http://www.mas.ecp.fr/vision/Personnel/iasonas/
Nikos Paragios[1,2]
http://www.mas.ecp.fr/vision/Personnel/nikos/

[1] Center for Visual Computing
Ecole Centrale Paris, France
[2] Equipe GALEN, INRIA Saclay
Île de France, Orsay, France
[3] Microsoft Research
Paris, France
[4] University of Crete
Grece

Large scale urban modeling has become a core component for a number of domains like the game industry, urban planning, etc. In such a context, scalability, modularity, compactness and semantic context are to be reconciled. Shape grammars (SGs) provide a powerful structural representation that naturally encompasses these objectives. In this talk, we will first consider procedural techniques which rest on SGs to tackle the automatic generation of realistic yet virtual architectural scenes [5]. We will then turn our attention to the image-based modeling of buildings where structured approaches are still at their infancy [1].

After explaining how SGs can be used for automatically creating virtual environments, we will concentrate on a major challenge attached to this task. It concerns the trade-off between the expressive power of a grammar and the capability to restrict its language to architecturally consistent models only. In this perspective, we have proposed two natural tools referred as tags to finely control the grammar derivation, introducing respectively the possibility to couple and to differentiate the treatments of similar primitive instances.

Encouraged by this preliminary success of SGs, we have investigated the benefits of such representations for the image-based modeling of existing environments. In [2, 3], we have introduced a link between the symbolic representation derived from the grammar and the statistical properties attached to semantic classes in the images. This link can be achieved in a generic way thanks to supervised learning techniques such as Randomized Forests, and is expressed in practice through posterior distributions $p(c|f)$ evaluating the likelihood of associating a class c with a feature vector f.

The previous development naturally leads to the following energy that is to be minimized with respect to the procedural semantic layout π.

$$E(\pi) = -\sum_{x\in\Omega} \log p(c_x^{\pi}|f_x) \qquad (1)$$

where Ω is the image domain, c_x^{π} is the semantic class predicted by π at pixel x and f_x is a feature vector extracted at this location. Using a simple hill climbing search, we obtained promising classification performance demonstrating how our procedural prior yields sharp improvements in comparison to simpler priors.

However, such a simple optimization strategy is prone to convergence failures and requires a large number of tested layouts before reaching the optimal one. To reduce the computational time, we explored in [4] an alternative strategy based on reinforcement learning (RL). In this context, the energy is decomposed as a sum of successive costs obtained at each step of the grammar derivation which can thus be interpreted as Markov decision process (MDP). In addition, the classical Q-Learning algorithm could be adapted to integrate data-driven guidance. This extension results in 50fold speedups over [3], while giving competitive labeling accuracy.

This approach is nonetheless limited to 2D procedural interpretations of a single image and assumes a binary split grammar - a special form that still leads to the same generative power as the general form. Therefore, to handle more complex scenarios including the recovery of the true 3D structure of a building, we are currently developing a second extension where multiple views of the building are captured. The previous energy can be extended to this new context and is completed by a concurrent objective function derived from a classical multi-view reconstruction. We obtained a multi-objective formulation where evolutionary algorithms (EAs) provide an adapted way to estimate the so-called Pareto front. Among the whole set of non-dominated solutions recovered in this way, we could extract automatically one that achieves a meaningful consensus between the two objectives. Reconstructions obtained with this approach were evaluated successfully with respect to classification and geometric accuracy. Subsidiary experiments conducted on the single view settings, demonstrated that the EA search is approximately as efficient as RL.

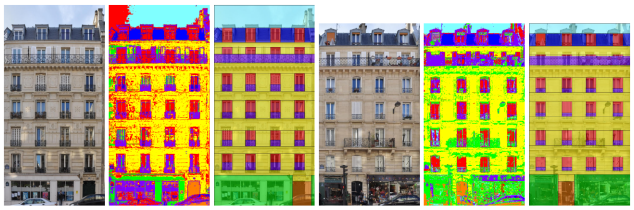

Figure 1: Parsing results obtained with RL - for each building, we depict the input image, an unstructured labeling and the procedural parsing.

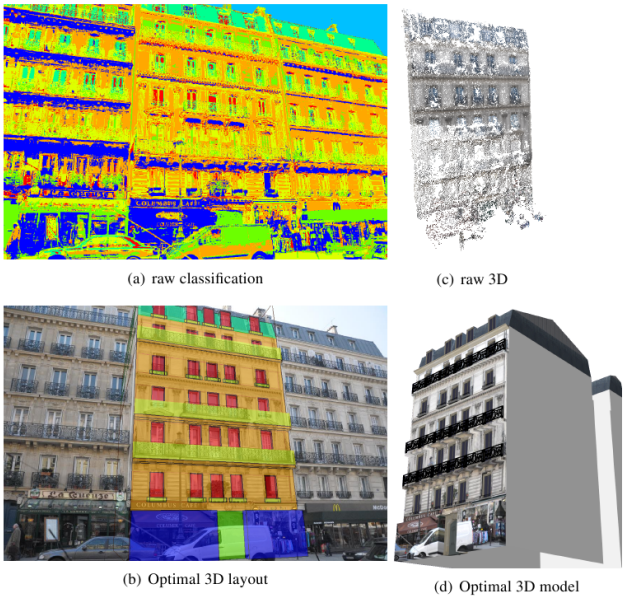

Figure 2: Results obtained with EAs in the multi-view settings.

[1] P. Müller, G. Zeng, P. Wonka, and L. Van Gool. Image-based procedural modeling of facades. In *ACM TOG*, 2007.

[2] L. Simon, O. Teboul, P. Koutsourakis, and N. Paragios. Random Exploration of the Procedural Space for Single-View 3D Modeling of Buildings. *IJCV*, 2011.

[3] O. Teboul, L. Simon, P. Koutsourakis, and N. Paragios. Segmentation of building facades using procedural shape priors. In *CVPR*, 2010.

[4] O. Teboul, I. Kokkinos, P. Koutsourakis, L. Simon, and N. Paragios. Shape Grammar Parsing via Reinforcement Learning. In *CVPR*, 2011.

[5] P. Wonka, M. Wimmer, F. Sillion, and W. Ribarsky. Instant architecture. In *ACM TOG*, 2003.