This CVPR2015 extended abstract is the Open Access version, provided by the Computer Vision Foundation.

A MRF Shape Prior for Facade Parsing with Occlusions

Mateusz Koziński, Raghudeep Gadde, Sergey Zagoruyko, Guillaume Obozinski and Renaud Marlet
Université Paris-Est, LIGM (UMR CNRS 8049), ENPC, F-77455 Marne-la-Vallée

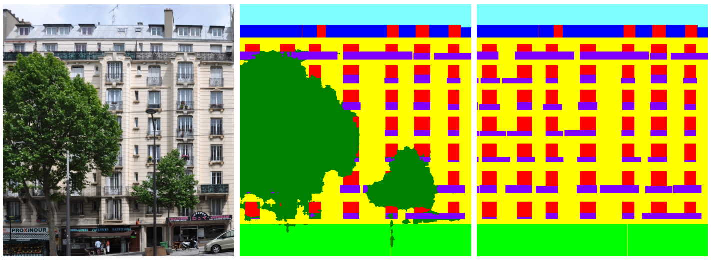


Figure 1: Example input image (left). Segmentation of the visible objects (middle) and facade structure (right) extracted by our algorithm.

The goal of facade parsing is to segment a rectified image of a building facade into regions corresponding to architectural elements, like windows, balconies and doors. The task differs from general-purpose image segmentation, because unlike general scenes, facades often contain a number of aligned elements and only some of their spatial combinations are semantically valid. For example, a balcony needs to be adjacent to the lower part of at least one window. In general, we consider a set of 'semantic constraints', specified by the user for a given dataset, that the segmentations have to satisfy. The quality of facade segmentation, as perceived by a human, suffers a lot if these semantic constraints are not respected.

Existing approaches to the problem can be divided into two categories. The frameworks with user-defined shape priors [4, 5] guarantee conformance of the resulting segmentations to the priors, but require a random exploration of the space of conforming segmentations for inference. The methods that impose structural constraints on results of general-purpose image segmentation algorithms [1, 3], do not offer a formal way of specifying a prior by the user, and often cannot guarantee conformance of the segmentation to the prior.

We propose a new method, that yields state-of-the-art results on a number of datasets, and advances the field in several directions, providing a unique combination of guaranteed satisfaction of user-defined priors, simultaneous alignment in 2D, modeling of shapes with irregular boundaries, simultaneous segmentation of occluding objects and occluded facade.

We propose a shape prior formalism called an 'adjacency pattern', a triple $A = (S,V,H)$ where S is a finite set of classes, and $V \subset S \times S$ and $H \subset S \times S$ are sets of ordered pairs of classes that can be assigned to vertically and horizontally adjacent pixels. A pair of vertically adjacent pixels can be labeled in such a way that a pixel of class s_1 is immediately below a pixel of class s_2 only if $(s_1,s_2) \in V$. The same holds for any pair of horizontally adjacent pixels and the set H. Adjacency patterns can be used to encode a number of interesting shape priors: grids (figure 2), grids with nonlinear boundaries (figure 3) and grids with monotonic boundaries (figure 4).

Encoding patterns with multiple misalignments by means of a single grid results in a number of classes that is exponential in the number of misalignments. To address this issue, we define a hierarchy of nested grids, where a cell of a coarse grid can be further subdivided into a finer grid. We propose to model such complex shapes using a hierarchy of adjacency patterns (figure 5). Transition between levels of the hierarchy is realized by mapping pixel classes of an adjacency pattern on a coarser level to other adjacency patterns on a finer level. We show in the paper that the hierarchical adjacency pattern can be represented as a standard, 'flat' adjacency pattern.

Finally, adjacency patterns enable explicit modeling of occlusions. Our algorithm simultaneously extracts the structure of the occluded facade and the boundary of the occluding object (figure 1).

The problem of finding the optimal segmentation conforming to an adjacency pattern can be formulated as the MAP-MRF problem over a 4-connected grid of pixels. We use the dual decomposition algorithm [2] to approximate the optimal segmentation. We decompose the problem into Markov chains over image rows and columns. A dedicated procedure for extracting the primal solution from statistics collected during the optimization routine guarantees that the resulting segmentation respects the constraints on classes of adjacent pixels.

Given the same data term, our algorithm outperforms existing approaches in terms of segmentation accuracy on a number of datasets.

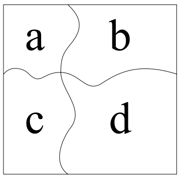
This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

a	b	a
c	d	c
a	b	a

H ↱	a	b	c	d
a	+	+	–	–
b	+	+	–	–
c	–	–	+	+
d	–	–	+	+

V ↱	a	b	c	d
a	+	–	+	–
b	–	+	–	+
c	+	–	+	–
d	–	+	–	+

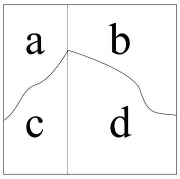
Figure 2: A repeating pattern with straight, axis-aligned boundaries. The tables represent allowed (+) and forbidden (–) pairs of horizontally (H) and vertically (V) adjacent pixels.



H ↱	a	b	c	d
a	+	+	+	–
b	–	+	–	+
c	+	–	+	+
d	–	+	–	+

V ↱	a	b	c	d
a	+	+	–	–
b	+	+	–	–
c	+	–	+	+
d	–	+	+	+

Figure 3: A non-repeating pattern with winding, axis-driven boundaries.



H ↱	a	b	c	d
a	+	+	+	–
b	–	+	–	–
c	–	–	+	+
d	–	–	–	+

V ↱	a	b	c	d
a	+	–	–	–
b	–	+	–	–
c	+	–	+	–
d	–	+	–	+

Figure 4: A non-repeating pattern on grid with monotonic boundaries.

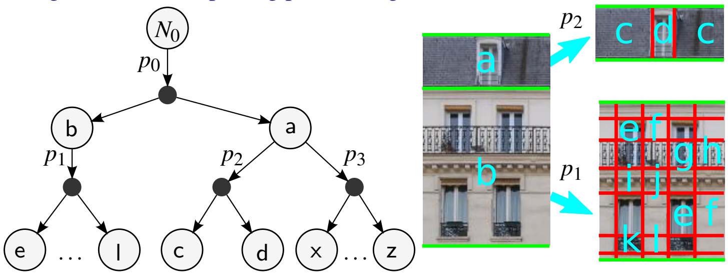


Figure 5: A hierarchy of adjacency patterns and a conforming segmentation. Gray nodes correspond to pixel classes. Black nodes correspond to adjacency patterns. Productions are marked next to arrows that map pixel classes to adjacency patterns. The hierarchy on the left encodes an alternative between production p_2 and production p_3 (not represented on the right).

[1] M. Pollefeys A. Cohen, A. Schwing. Efficient structured parsing of facades using dynamic programming. In *CVPR*, 2014.

[2] N. Komodakis, N. Paragios, and G. Tziritas. MRF energy minimization and beyond via dual decomposition. *PAMI*, 33(3), 2011.

[3] A. Martinovic, M. Mathias, J. Weissenberg, and L. Van Gool. A three-layered approach to facade parsing. In *ECCV*, 2012.

[4] H. Riemenschneider, U. Krispel, W. Thaller, M. Donoser, S. Havemann, D. Fellner, and H. Bischof. Irregular lattices for complex shape grammar facade parsing. In *CVPR*, 2012.

[5] O. Teboul, I. Kokkinos, L. Simon, P. Koutsourakis, and N. Paragios. Shape grammar parsing via reinforcement learning. In *CVPR*, 2011.