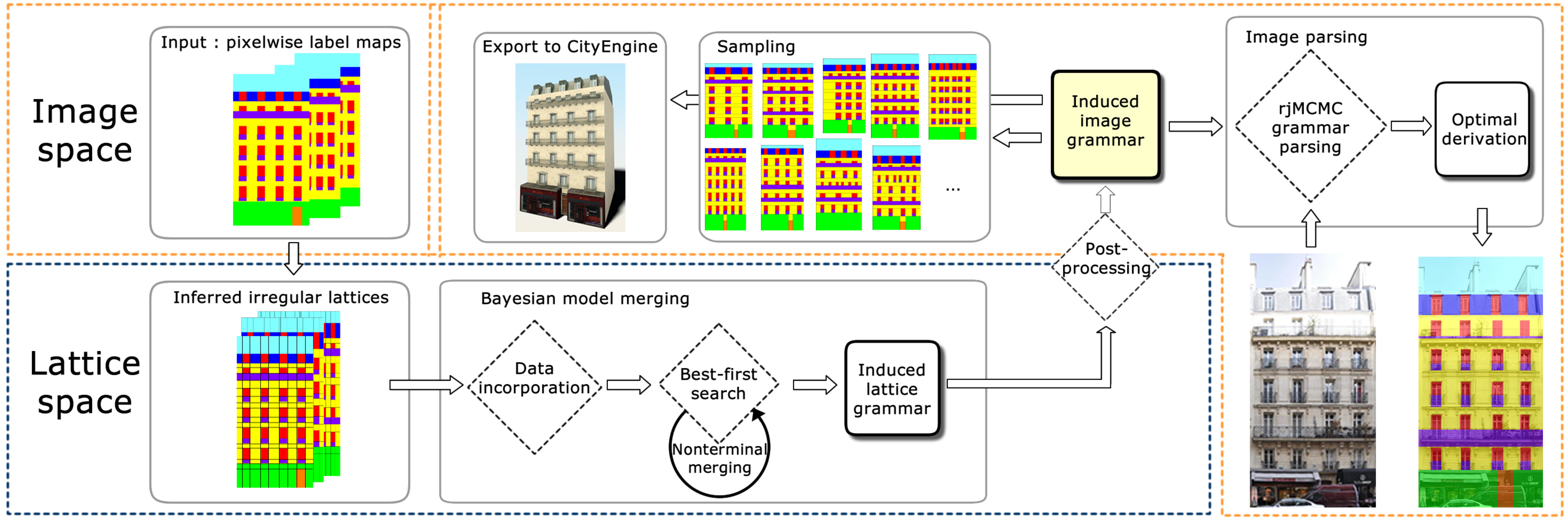
Bayesian Grammar Learning for Inverse Procedural Modeling

Anđelo Martinović[1], Luc Van Gool[1,2]

[1]ESAT/PSI/VISICS, KU Leuven [2]Computer Vision Lab, ETH Zurich

Method overview



Introduction

Shape grammars are a powerful tool for synthesizing novel building designs and reconstructing existing buildings. So far, a human expert was required to write grammars for specific building styles. We present an approach to automatically learn procedural split grammars from labeled building facades. The learned grammar can be used to:

1. **Create** novel instances of the same building style
2. **Parse** existing facade imagery, outperforming bottom-up classifiers, on-par with approaches that use a manually designed grammar

2D-ASCFGs

2-Dimensional Attributed Stochastic Context-Free

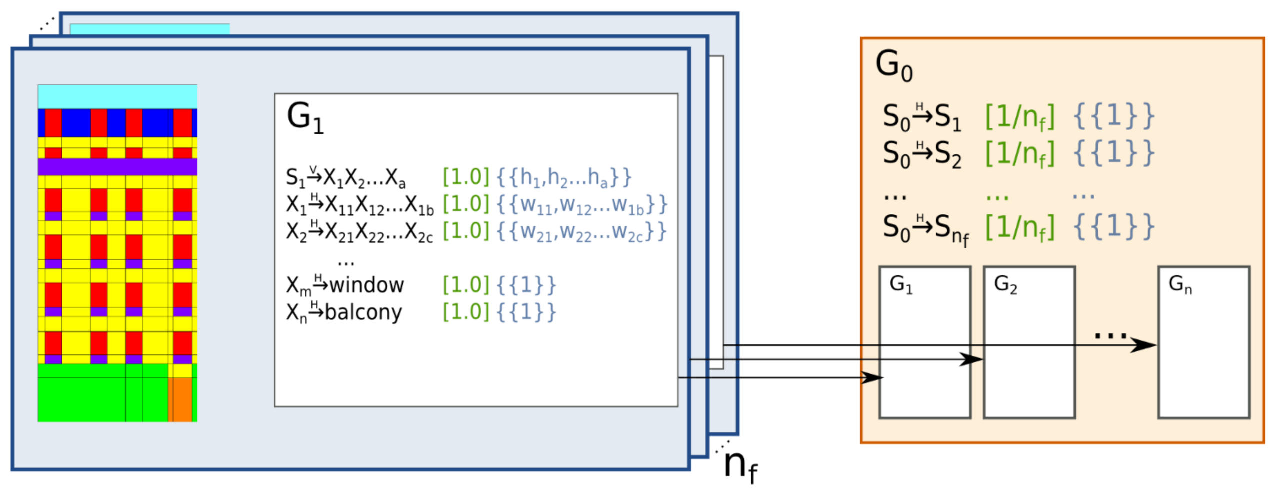
2D-ASCFG

G = (N, T, S, R, P, A)

Nonterminal symbols, Terminal symbols, Axiom, Production rules, Rule probabilities, Rule attributes

Data incorporation

Create a stochastic grammar that generates only the input lattices:



Bayesian model merging

Search for the grammar model that yields the best trade-off between the fit to the input data and a general preference for simpler models [2]

$$minimize\ E(G|D) = -w\, log\, P(G) - log\, P(D|G)$$

Prior

$$P(G) = P(G_s)P(\theta_g|G_s)$$

Structure prior
MDL principle

Parameter prior
Symmetrical Dirichlet

Likelihood

- ML estimate of $P(D|G_s)$ using Viterbi assumption
- Input data parsed using 2D Earley parser [1]
- Optimal values of θ_g estimated using EM

$$\hat{P}(X \to \lambda) = \frac{\hat{c}(X \to \lambda)}{\sum_\mu \hat{c}(X \to \mu)}$$

Search in model space

```
CurrentGrammar ← InitialGrammar
while 1 do
    Candidates ← ModelMerging(CurrentGrammar)
    BestCandidate ← argmin(E(Candidates))
    if E(BestCandidate) < E(CurrentGrammar) then
        CurrentGrammar ← BestCandidate
    else
        return CurrentGrammar
```

Model merging

$$Z_1 \to \mu_1 X_1 \lambda_1 \quad \overset{merge}{\dashrightarrow} \quad Z_1 \to \mu_1 Y \lambda_1$$
$$Z_2 \to \mu_2 X_2 \lambda_2 \qquad\qquad Z_2 \to \mu_2 Y \lambda_2$$

$$X_1 \to \lambda_1 \quad \overset{merge}{\dashrightarrow} \quad Y \to \lambda_1$$
$$X_2 \to \lambda_2 \qquad\qquad Y \to \lambda_2$$

Final model creation

Collapse sequences of the same non-terminal symbol in a production to a single symbol with correspondingly modified attributes:

$$X \to \lambda Y Y \mu \quad \overset{collapse}{\dashrightarrow} \quad X \to \lambda Y \mu$$
$$A = \{\{s_1, y_1, y_2, s_2\}\} \qquad A = \{\{s_1, y_1 + y_2, s_2\}\}$$

For every production $p = (X \to \lambda_1 \ldots \lambda_k)$, fit a $(k-1)$-variate Gaussian distribution $\phi(A) = \mathcal{N}(\bar{\mu}, \hat{\Sigma})$ to the set of its attributes $A(p) = \{\alpha_1 \ldots \alpha_n\}$

$$\bar{\mu} = \frac{1}{n}\sum_{j=1}^{n} \alpha_j$$

$$\hat{\Sigma} = \frac{1}{n-1}\sum_{j=1}^{n}(\alpha_j - \bar{\mu})(\alpha_j - \bar{\mu})^T$$

Grammar size

	Initial grammar G_0	Induced, $w = 0.3$	Induced, $w = 1.0$
$\lvert N \rvert$	126.8 ± 6.61	26.6 ± 0.89	14 ± 0.0
$\lvert R_h \rvert$	121.8 ± 6.61	65 ± 6.70	27.8 ± 2.68
$\lvert R_v \rvert$	33 ± 0.0	15.6 ± 2.60	11 ± 1.41

Image space parsing

Similar to [3], find the optimal derivation using rjMCMC with the chain energy defined as:

$$E(\delta|I) = -\sum_{s\in\delta} log\, P(r_s) - \sum_{s\in\delta} log\, \phi(A(r_s)) - log\, P(I|\delta)$$

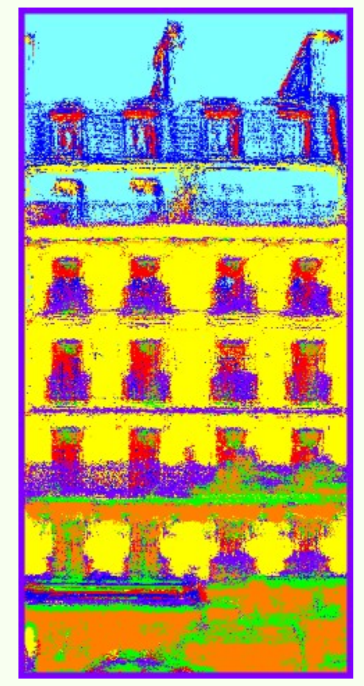
$$E_\delta = E_\delta^{rule} + E_\delta^{attribute} + E_\delta^{image}$$

Penalize low probability rules

Penalize deviations from learned attributes

Random Forest pixel classifier

$$E^{image} = \sum_{t\in\delta}\sum_{x_i \in t} -log\, P_{RF}(l_t|x_i)$$



Diffusion move

$$\rho_{\delta\to\delta'} = min\{1, \frac{p(\delta'|I)}{p(\delta|I)}\} = min\{1, e^{-(E_{\delta'} - E_\delta)}\}$$

Jump move

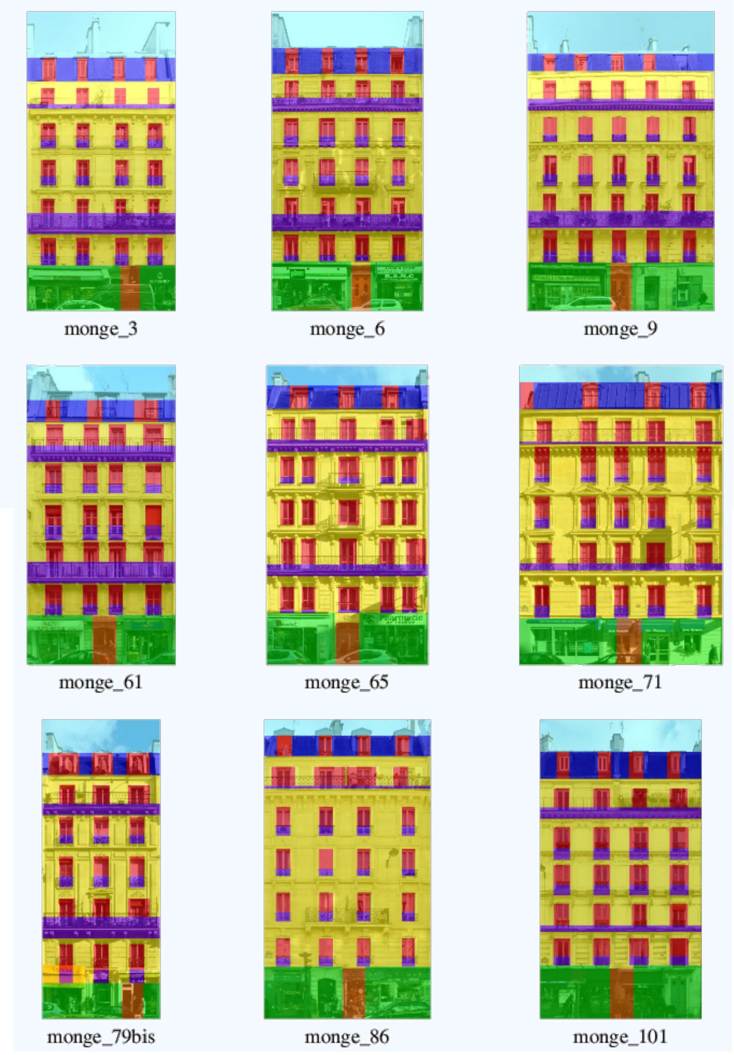
$$\rho_{\delta\to\delta'} = min\{1, \frac{q_{\tau'}(h)}{q_\tau(h)} e^{-[(E_{\delta'}^{img} + E_{\delta'}^{attr}) - (E_\delta^{img} + E_\delta^{attr})]}\}$$

Results

Class	RF[4]	RL[5]	Ours
Window	29	62	**66**
Wall	58	82	**80**
Balcony	35	58	**49**
Door	79	47	**50**
Roof	51	66	**71**
Sky	73	95	**91**
Shop	20	88	**81**
Overall	48.55	74.71	**74.82**

All of the above methods use pixel-based bottom-up information.

Facade parsing



Sampling novel buildings

Samples from the Bayes-optimal grammar

High prior weight Grammar is over-generalizing

References

[1] A. Martinović and L. Van Gool. Earley parsing for 2D stochastic context free grammars. Technical Report KUL/ESAT/PSI/1301, KU Leuven, 2013.
[2] A. Stolcke. Bayesian Learning of Probabilistic Language Models. PhD thesis, UC Berkeley, 1994.
[3] J. O. Talton, Y. Lou, S. Lesser, J. Duke, R. Mech, and V. Koltun. Metropolis procedural modeling. In SIGGRAPH, 30(2), 2011.
[4] O. Teboul, L. Simon, P. Koutsourakis, and N. Paragios. Segmentation of building facades using procedural shape priors. In CVPR, 2010.
[5] O. Teboul, I. Kokkinos, L. Simon, P. Koutsourakis, and N. Paragios. Shape grammar parsing via reinforcement learning. In CVPR, 2011.