Hierarchical Co-Segmentation of Building Facades

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

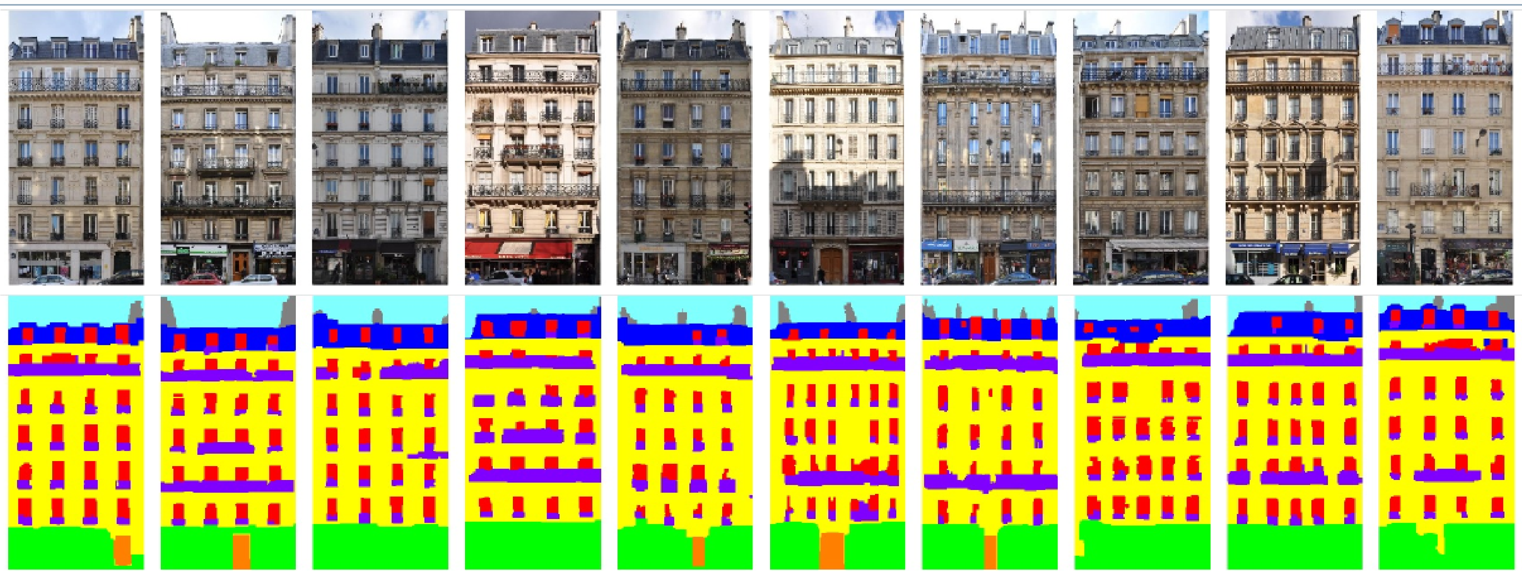
[1]ESAT/PSI/VISICS, KU Leuven [2]Computer Vision Lab, ETH Zurich

Motivation

Previous approaches for **facade structure learning** analyse each facade in isolation, and depend on user interaction or ground-truth labelings, e.g. boxes or pixelwise annotations.
We propose the following:

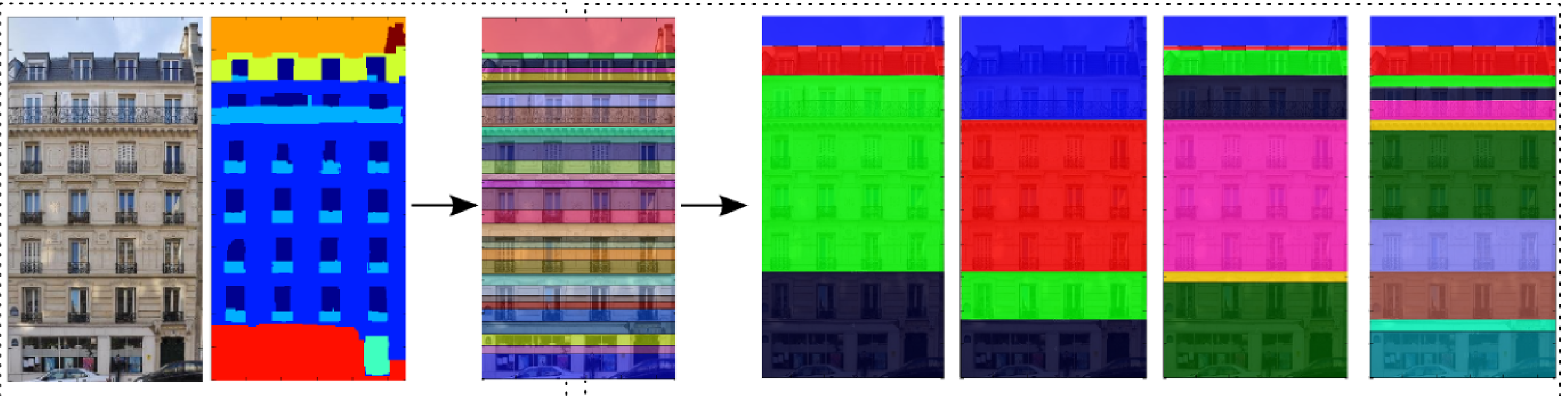
1. Augment the input images with their **semantic labeling** [1];
2. Utilize **co-segmentation** [2] to create consistent axis-aligned segmentations;
3. Create **hierarchical decompositions** as high-level structural representations of facades.

Our method captures essential structural information, useful for facade retrieval or virtual facade synthesis with induced shape grammars.



Our method

1. Segment proposals



- Oversegmentation in K slices
- Splitting function

$$\Upsilon(z) = \Upsilon_{IG}(\mathbf{F}_z) \cdot \Upsilon_{IC}(\mathbf{F}_z) \cdot \Upsilon_{LB}(\mathbf{L}_z) \cdot \Upsilon_{LC}(\mathbf{L}_z)$$

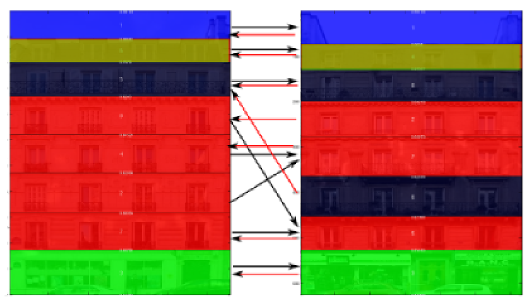
Image gradients Image content Label gradients Label content

- Peak detection with binary search

- Generate many **random segmentations** by varying the number of segments
  1. Select random cluster centers
  2. Assign slices to segments based on a distance measure
  3. Iterate using K-medoids (Expectation-maximization)
- Reorder the segments based on their frequency of appearance w_s
- Keep n best segments

2. Co-segmentation

Pairwise

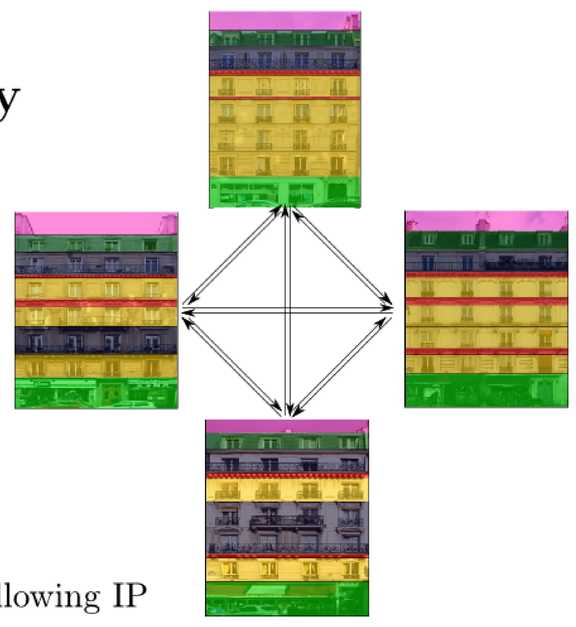


- Maximize the score function by finding the optimal:
  - Set of segments $\mathbf{x}_i$ for each image i
  - Many-to-one mappings $\mathbf{y}_{ij}$, i.e. correspondences between segments in images i and j

$$\max \underbrace{\sum_{i \in \{1,2\}} \mathbf{x}_i^T \mathbf{w}_i^{seg}}_{\text{Unary term}} + \lambda \underbrace{\sum_{ij \in \{12,21\}} \mathbf{y}_{ij}^T \mathbf{w}_{ij}^{cor}}_{\text{Pairwise term}}$$

Segment indicator variables; Segment goodness weighted by size; Mapping indicator variables; Similarity of two segments

Multiway



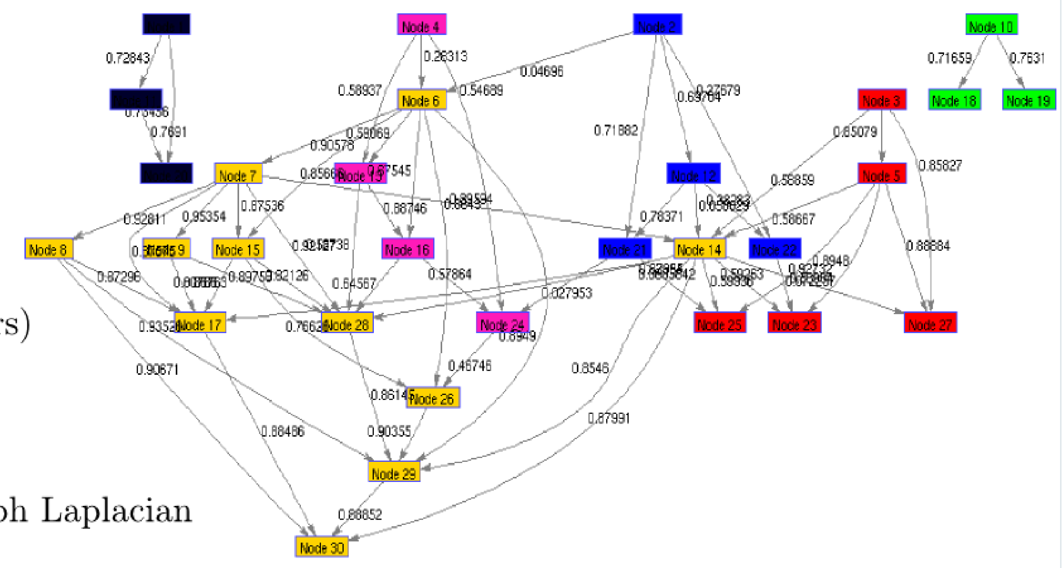
- Solve the following IP

$$\max \sum_{i=1}^{N} \mathbf{x}_i^T \mathbf{w}_i^{seg} + \frac{\lambda}{N-1} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \mathbf{y}_{ij}^T \mathbf{w}_{ij}^{cor}$$

- Subject to
  - Cover constraints
  - Mapping constraints
  - Integrality constraints (relaxed for y variables)

3. Segment clustering

- Multiway co-segmentation result: **directed co-segmentation graph**
  - Nodes: all selected segments
  - Edges: all optimal mappings
- Spectral clustering discovers natural clusters in the data (node colors)
- Number of clusters κ determined using the eigengap heuristic
  - Sort the eigenvalues λ_i of the graph Laplacian (in ascending order)
  - Pick κ as $argmax_i(\lambda_{i+1} - \lambda_i)$.



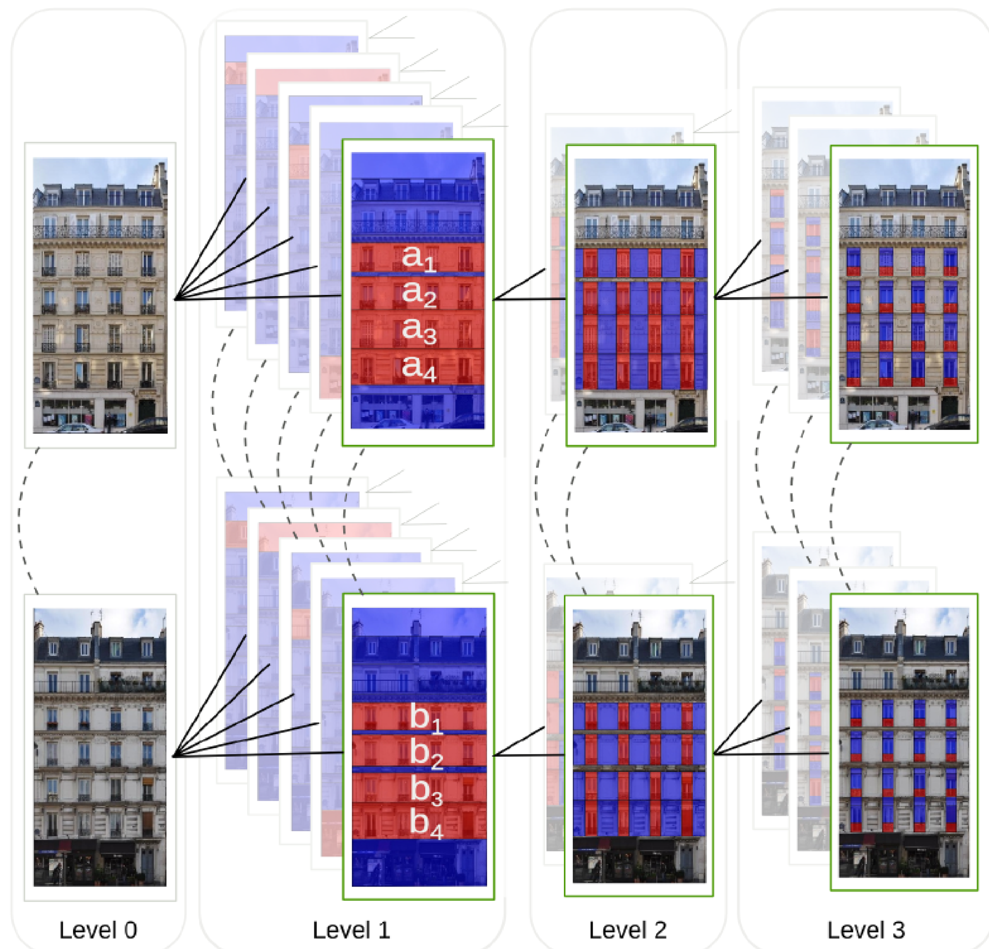
4. Hierarchy creation

- Perform the co-segmentation **recursively** for each of the detected clusters (e.g. all floors, all shops...)
- Problem: number of elements for co-segmentation increases dramatically (20 facades -> 80 floors -> 400 windows)
- Solution: segment synchronization
  - Data support averaging
  - Representative creation

$$\Upsilon^{avg} = \frac{1}{|\Psi|} \sum_{s \in \Psi} \Upsilon(s)$$

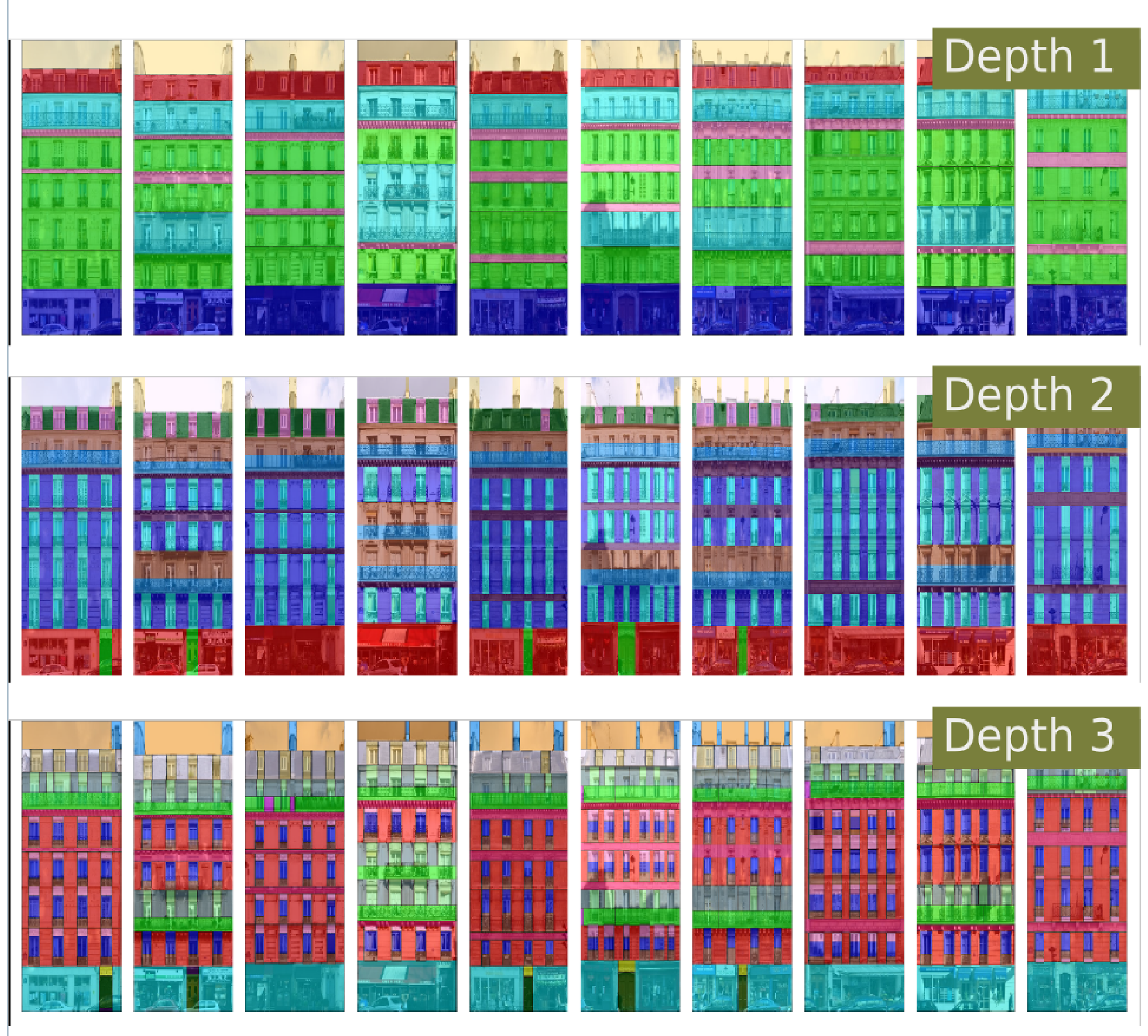
- Splitting direction determined **adaptively**
  - Co-segmentation performed in both directions
  - Select the direction which produces a more consistent result
  - Segmentation consistency based on scope similarity

$$w(z_i, z_j) = \mathbf{y}_{ij}^T \mathbf{w}_{ij}^{cor} + \mathbf{y}_{ji}^T \mathbf{w}_{ji}^{cor}, w \in [0, 2]$$



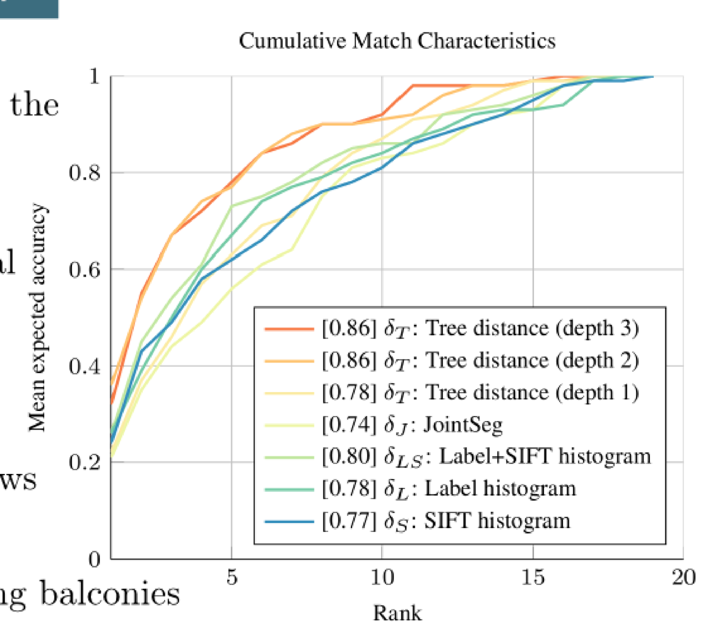
Results

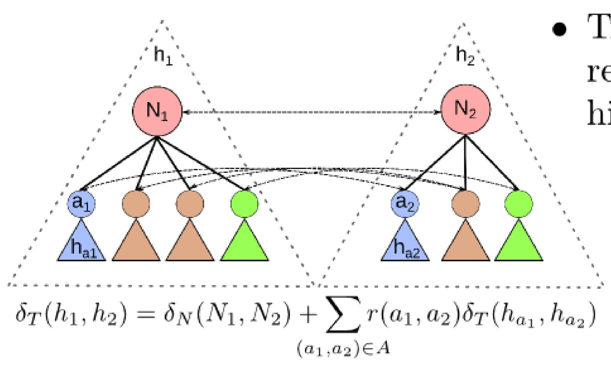
Qualitative



Facade retrieval

- Retrieval evaluated on the ECP dataset [3]
- Gold distance [4]: number of architectural changes between two facades
  - Number of floors
  - Number of windows
  - Door position
  - Position of running balconies





$$\delta_T(h_1, h_2) = \delta_N(N_1, N_2) + \sum_{(a_1, a_2) \in A} r(a_1, a_2) \delta_T(h_{a_1}, h_{a_2})$$

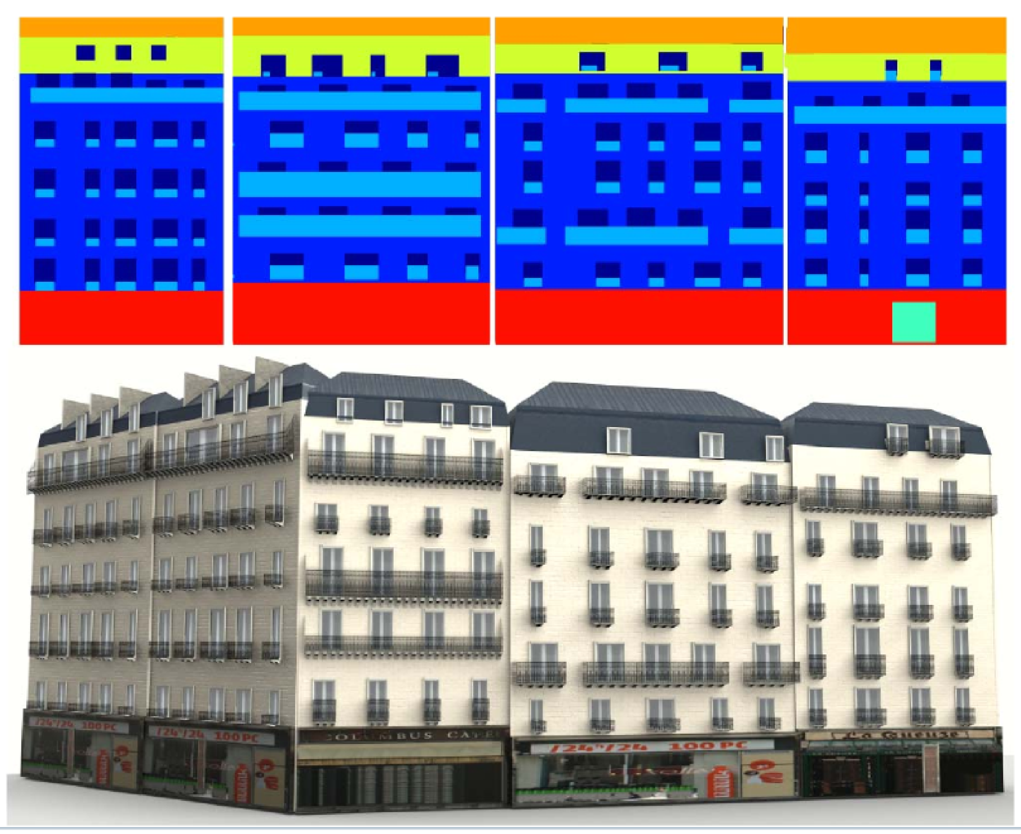
- Tree distance measure δ_T defined recursively between two induced hierarchical segmentations
  - Node distance calculated with the Smith-Waterman algorithm [5]
- Our tree distance measure provides top results
- Performance increases by using deeper hierarchies

Grammar induction

- Transforming hierarchies into procedural split rules of the form

$$X^\alpha \rightarrow split(dir)\{r_1^\alpha : b_1^\alpha | r_2^\alpha : b_2^\alpha ... | r_n^\alpha : b_n^\alpha\}$$

- Aggregating rules having the same structure (but different size vectors) into a single rule and fitting a multivariate Gaussian to the set of size vectors
- Sample from the induced grammars:



References

[1] A. Martinović, M. Mathias, J. Weissenberg and L. Van Gool. A three-layered approach to facade parsing. In ECCV, 2012.
[2] Q. Huang, V. Koltun and L. Guibas. Joint shape segmentation with linear programming. In SIGGRAPH, 30(6), 2011.
[3] O. Teboul, I. Kokkinos, L. Simon, P. Koutsourakis and N. Paragios. Parsing facades with shape grammars and reinforcement learning. In TPAMI, 35(7), 2013.
[4] J. Weissenberg, H. Riemenschneider, M. Prasad and L. Van Gool. Is there a procedural logic to architecture? In CVPR, 2013.
[5] T. Smith and M. Waterman. Identification of common molecular subsequences. In Journal of Molecular Biology, 147(1), 1981.

KU LEUVEN ETH zürich www.varcity.eu