Explicit Occlusion Modeling for 3D Object Class Representations

M. Zeeshan Zia† Michael Stark‡ Konrad Schindler†

† Photogrammetry and Remote Sensing, ETH Zurich

‡ Stanford University and Max Planck Institute for Informatics

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Motivation

Estimate 3D shape and pose even for partially occluded object instances in monocular images.

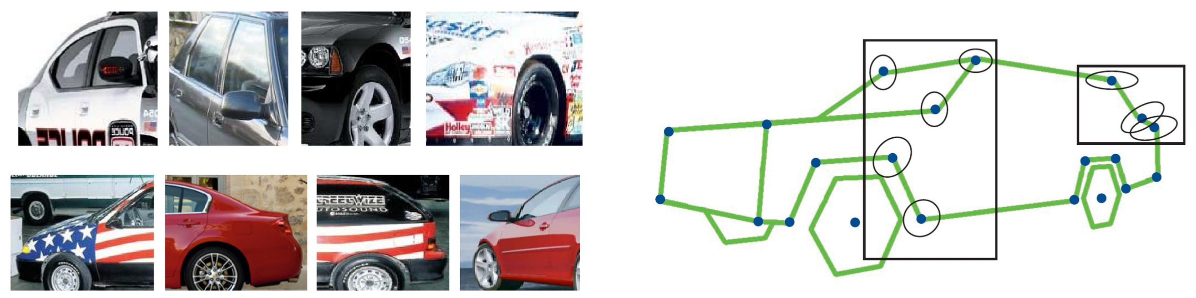
Contributions

- Explicit occluder representation for detailed 3D object class models.
- Complete framework for detection and reconstruction based on proven building blocks.
- 3D reasoning tightly coupled with 2D appearance matching.

Data set and source code

- Code, data, annotations being made public http://www.igp.ethz.ch/photogrammetry/downloads

Multi-layer architecture



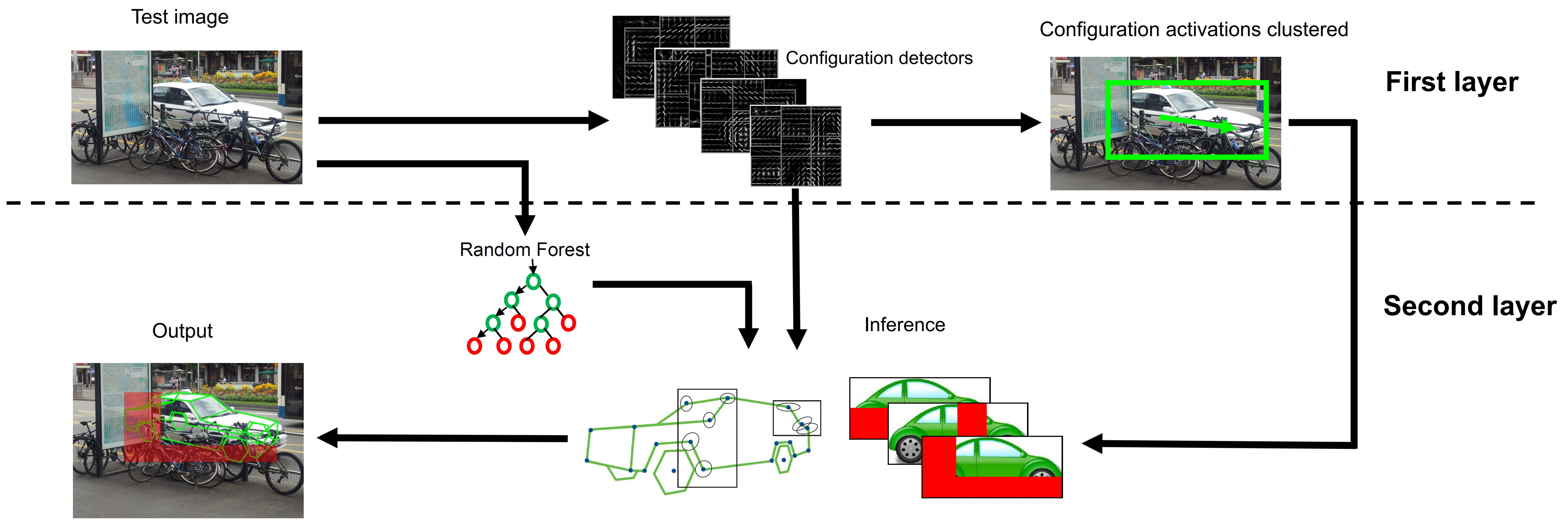
First layer: localize objects coarsely in 2D

- Parts: local windows centered at wireframe vertices
- Spatially contiguous sets of parts called *part configurations*.
- Single component DPM detector trained for each part configuration (118 detectors in our tests)
- Part configuration detections vote for full object bounding box, coarse pose, and part locations.

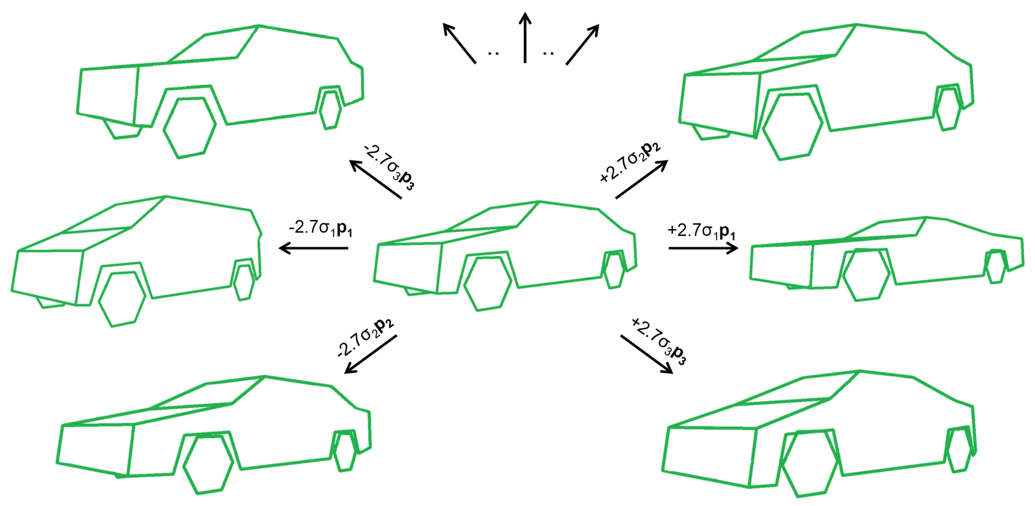
Second layer: detailed 3D reasoning

- Random forest based part detection
- Deformable model matching, occluder reasoning

Overview

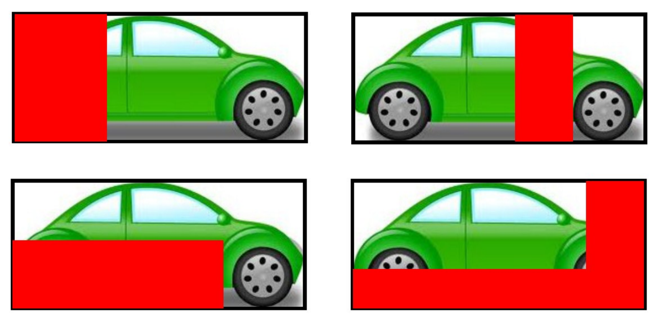


Geometric model



- Deformable 3D wireframe model
- Trained on 3D CAD data

Explicit occluder representation



- Enumerate exhaustive set of discrete occluder masks (288 masks in our tests)
- Block the view onto a spatially connected region of the object
- Neighborhood between masks: rank order w.r.t. Hamming distance
- Sample masks and set part visibility accordingly

Objective function formulation

$$\mathcal{L}(\mathbf{h}) = \max_{\varsigma} \left[\frac{1}{\sum_{j=1}^{m} o_j(\mathbf{s},\boldsymbol{\theta},a_0)} \sum_{j=1}^{m} \left(\mathcal{L}_v + \mathcal{L}_o + \mathcal{L}_c \right) \right]$$

where,

$$\mathcal{L}_v = o_j(\mathbf{s},\boldsymbol{\theta},a) \log \frac{S_j(\varsigma,\mathbf{x}_j)}{S_b(\varsigma,\mathbf{x}_j)} \,,$$

$$\mathcal{L}_o = \left(o_j(\mathbf{s},\boldsymbol{\theta},a_0) - o_j(\mathbf{s},\boldsymbol{\theta},a) \right) c \,,$$

$$\mathcal{L}_c = \frac{o_j(\mathbf{s},\boldsymbol{\theta},a)}{p} \sum_{i=1}^{p} v_{ij} \log \left(1 + \lambda \mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu}_{ij}, \boldsymbol{\sigma}_{ij}^2) \right) .$$

- $\mathcal{L}_v$: detection scores for the visible parts,
- $\mathcal{L}_o$: fixed likelihood for parts occluded by mask,
- $\mathcal{L}_c$: agreement of parts with detected configurations.
- o_j : hidden occlusion state given shape, pose, and occlusion mask.

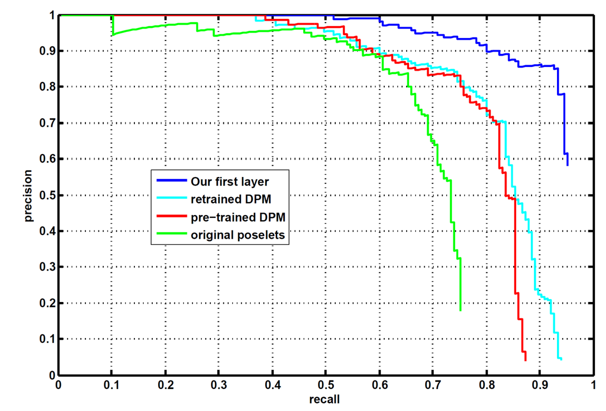
Inference

- Model-driven, smoothing-based optimization [Leordeanu&Hebert, 2008]
- Start from multiple randomly perturbed initializations, maintain multiple hypotheses.

Results

Example detections using our full system

Occlusion-agnostic model (l) vs. our full system (r)



Object detection accuracy of different 2D detectors

	Full dataset	< 80% visibility	< 60% visibility
Total cars	165	96	48
Detected	147	85	42

First-layer detection results (bounding box and 1D pose). Subsequent second-layer results are given for detected instances.

	Full dataset	< 80% visibility	< 60% visibility
avg shape in 2D bounding box	-	-	-
occlusion-agnostic 3D model	79.5%	76.7%	75.6%
w/o configurations (ours)	84.4%	82.6%	80.1%
w/ configurations (ours)	**85.6%**	**84.7%**	**83.1%**

Part-level occlusion prediction (percent correctly classified parts)

	Full dataset	< 80% visibility	< 60% visibility
avg shape in 2D bounding box	32.0%	33.6%	39.7%
occlusion-agnostic 3D model	80.0%	75.6%	74.5%
w/o configurations (ours)	82.5%	80.0%	79.8%
w/ configurations (ours)	**82.7%**	**80.7%**	**83.5%**

Part localization accuracy (percent correctly localized parts)