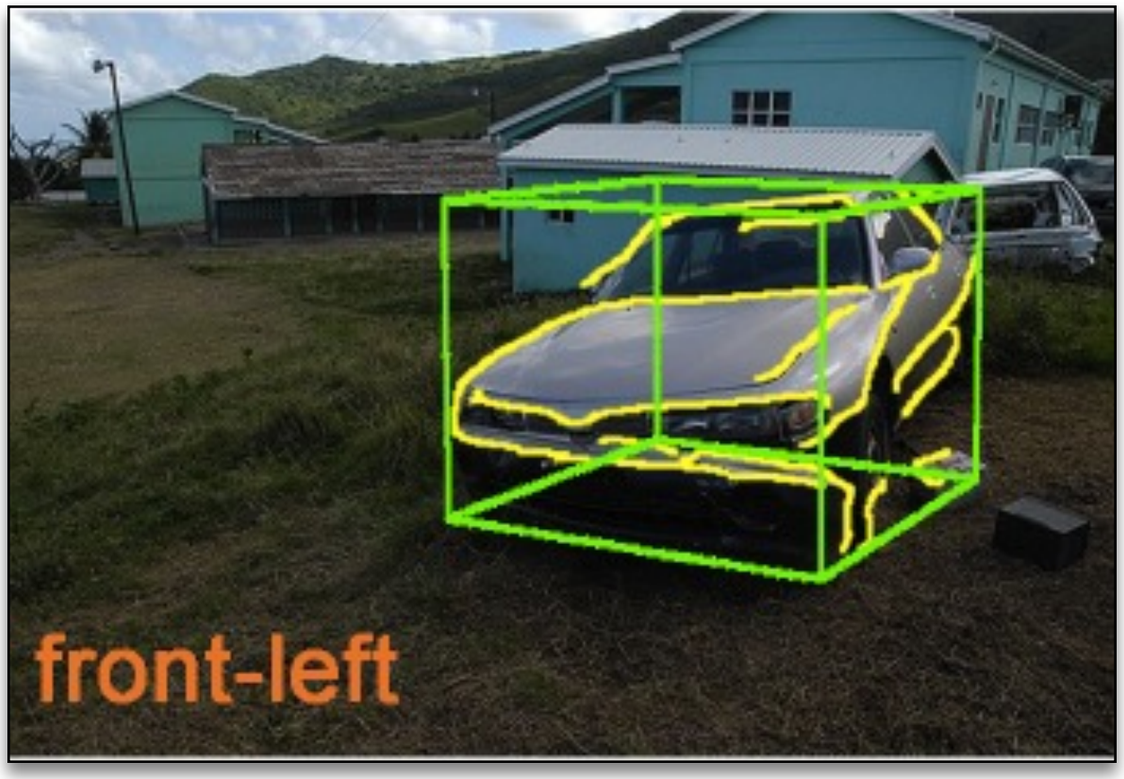
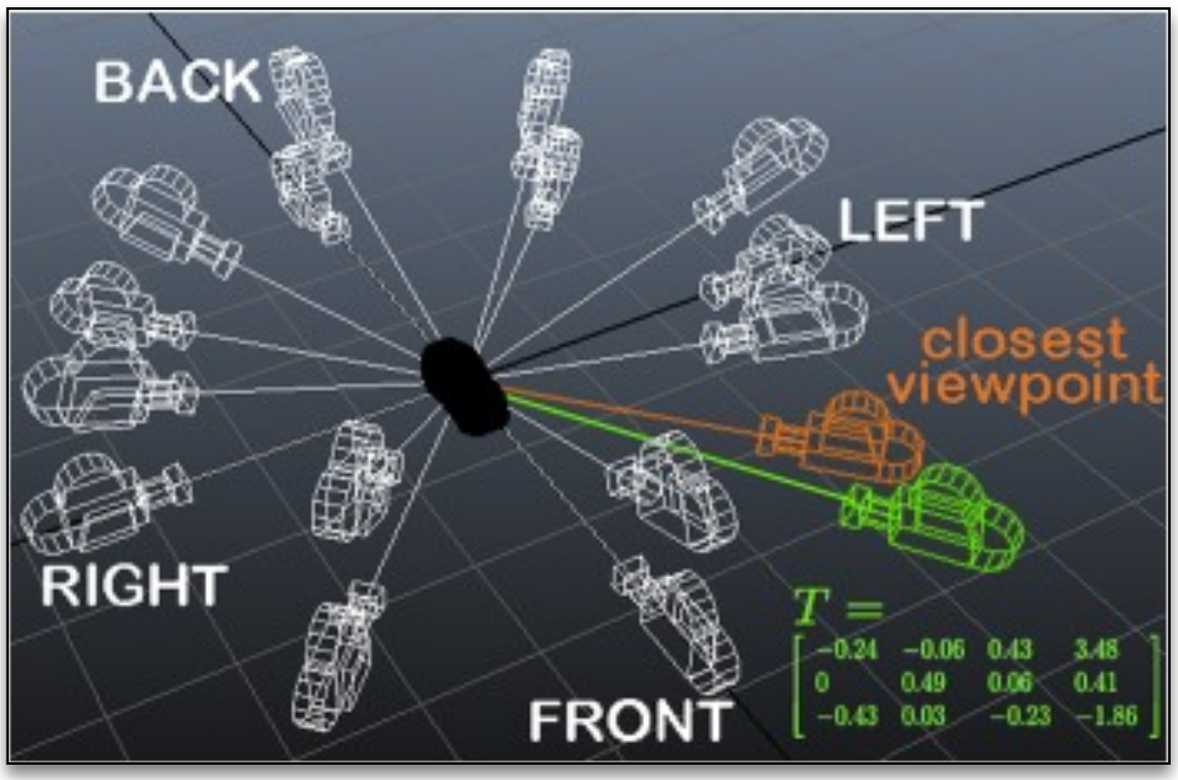
From Contours to 3D Object Detection and Pose Estimation

Nadia Payet and Sinisa Todorovic

Oregon State University

Problem Statement





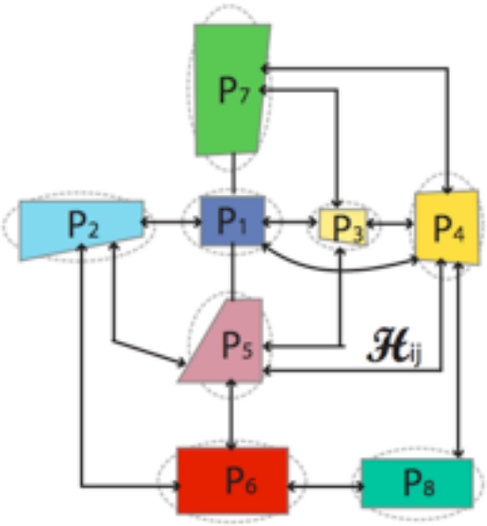
Given a single image:

1. Detect an object of interest
2. Delineate its boundaries
3. Estimate its continuous 3D pose

Prior Work

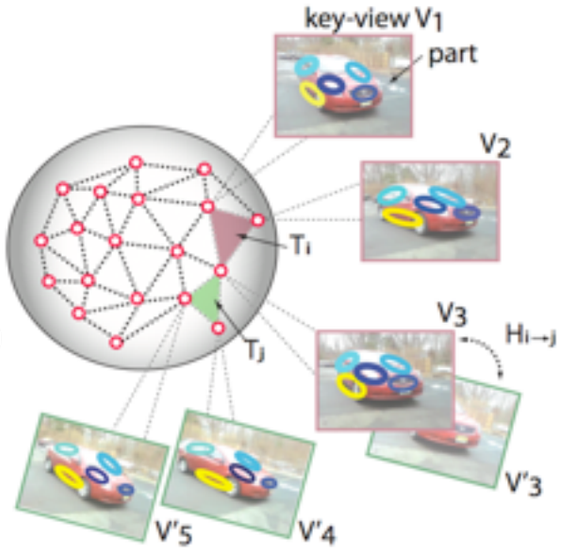
Generative models
e.g., aspect graphs

Koendrik & Doorn 79
Kushal et al. 04
Saverese & Fei-Fei 07-09
Arie & Basri 09
Hu & Zhu 10



Discriminative models
e.g., structured prediction

Hoiem et al. 07
Su et al. ICCV 09
Ozuysal et al. 09
Liebelt & Schmid 08-10
Gu & Ren 10

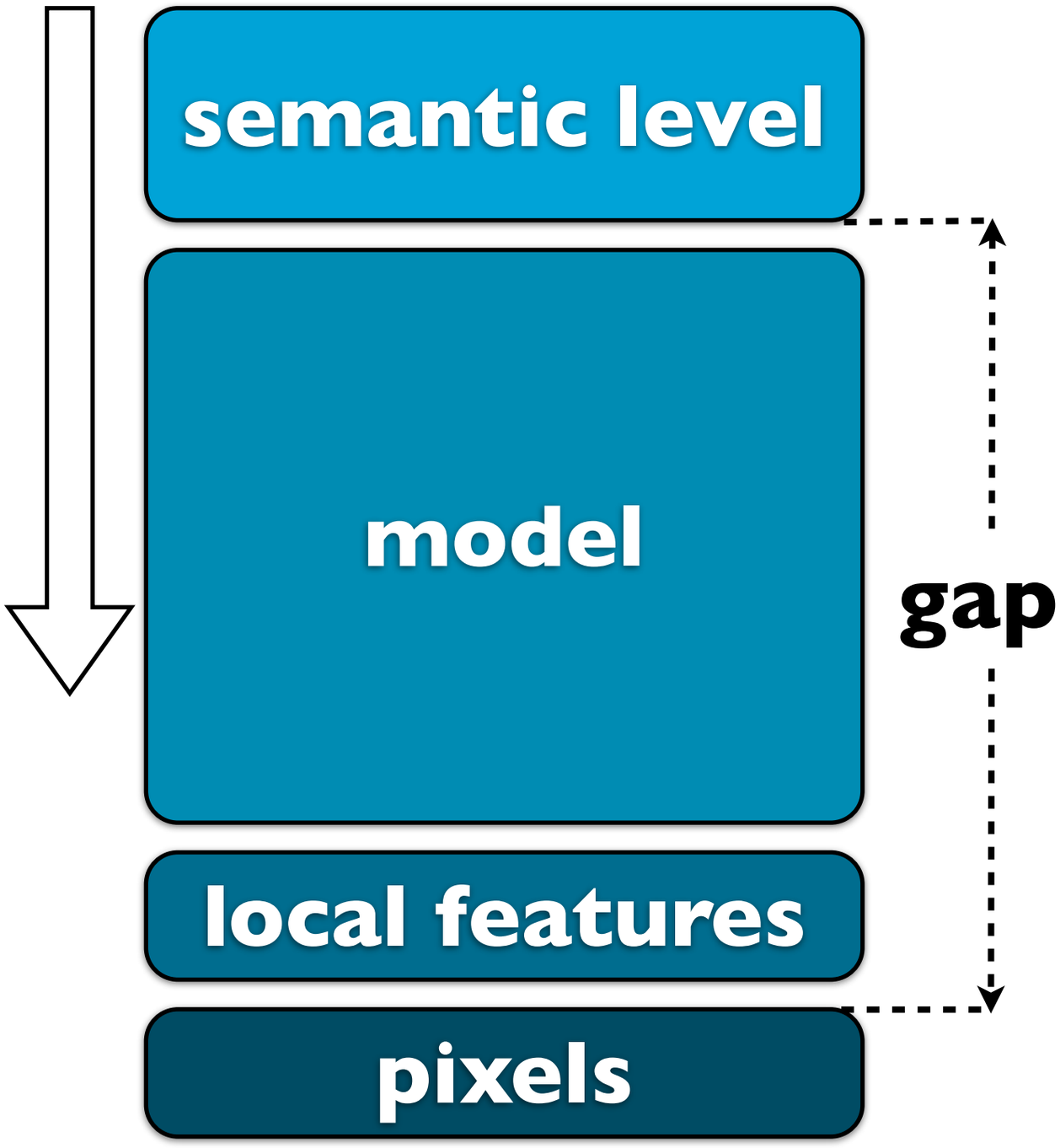


Main characteristics of recent work:

- Local image features
- Sophisticated models
- 3D pose = Interpolation of viewpoint classes

To Bridge the Semantic Gap...

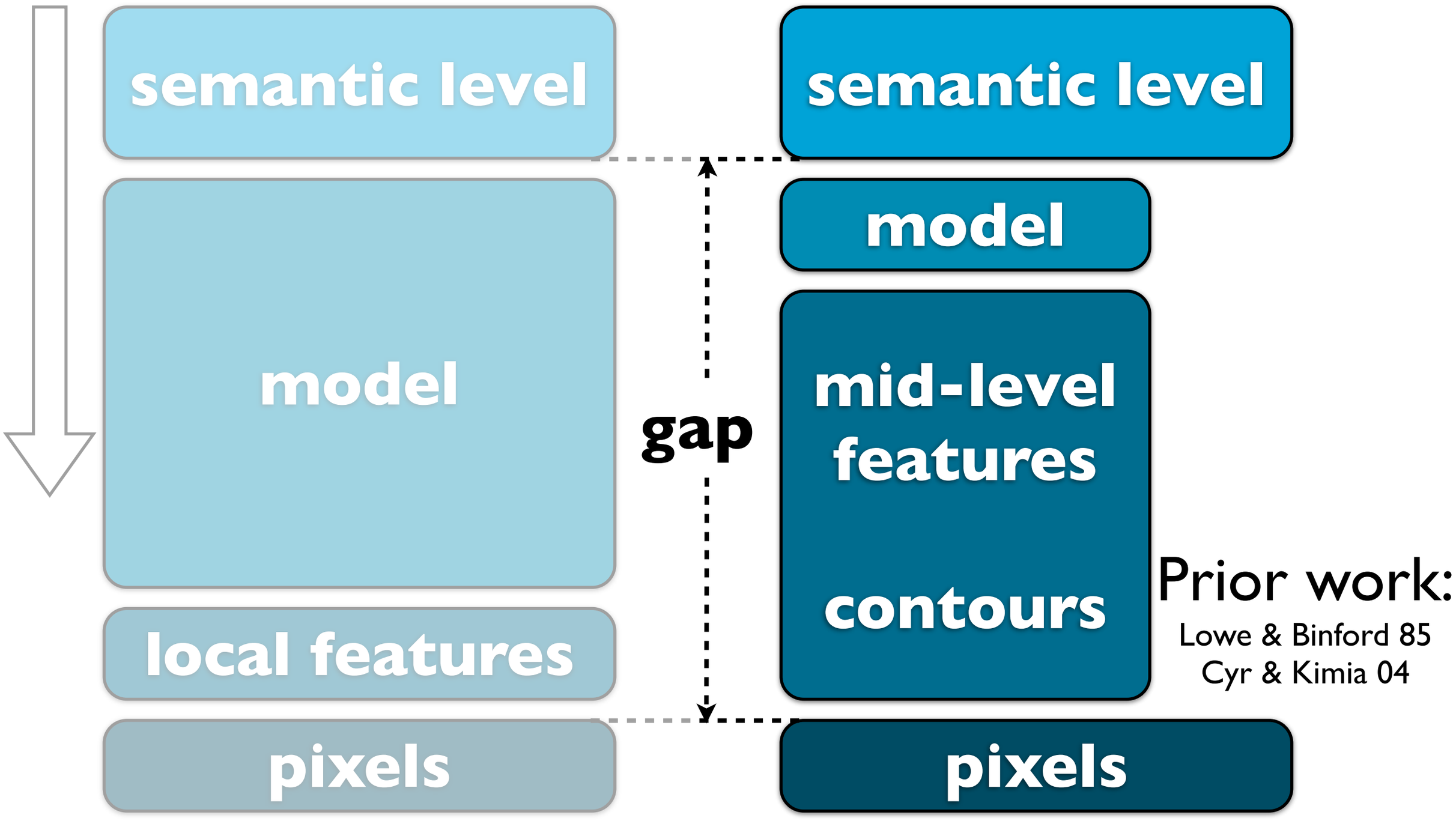
Recent work, typically



To Bridge the Semantic Gap...

Recent work, typically

Our approach



Prior work:

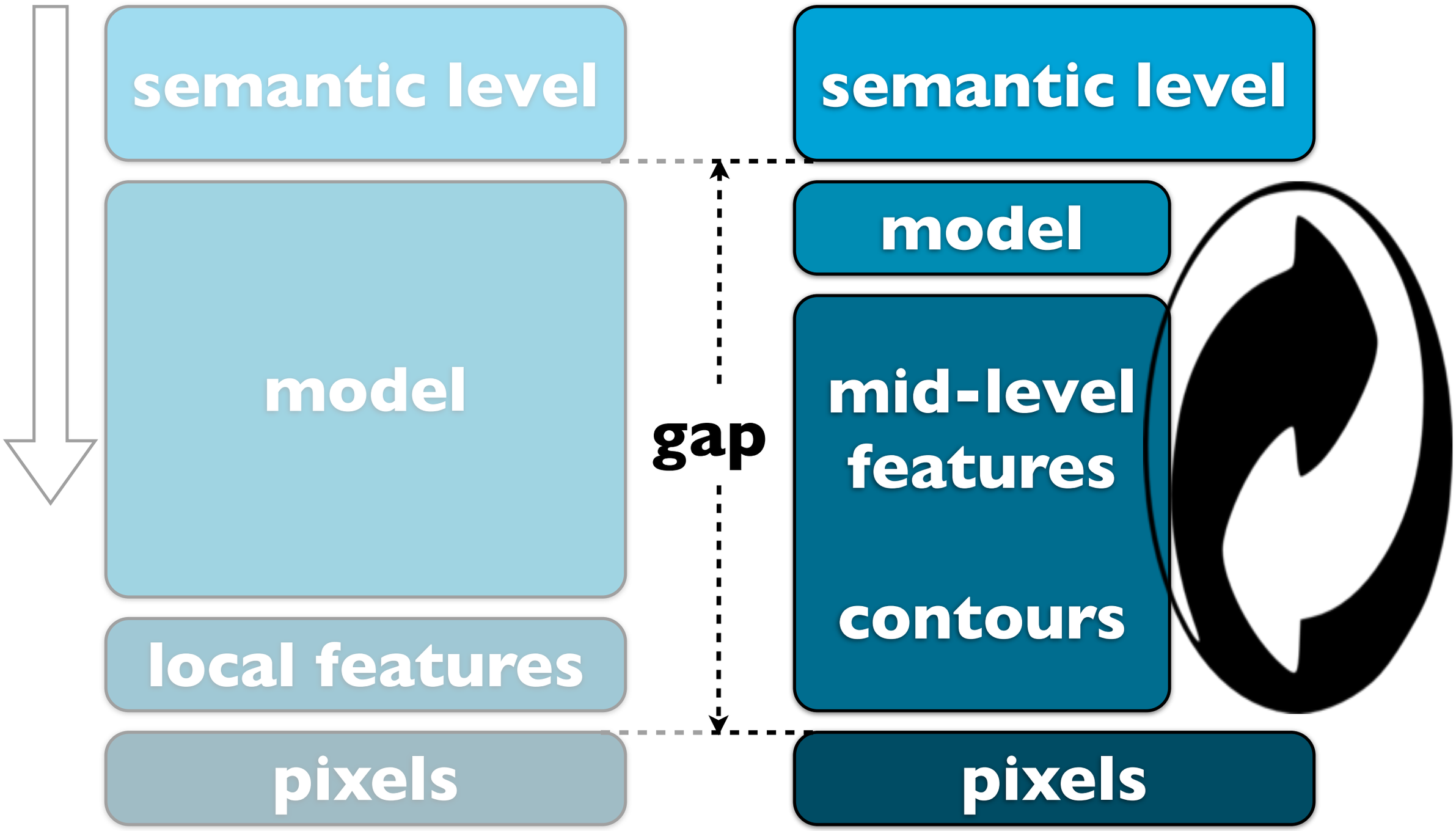
Lowe & Binford 85

Cyr & Kimia 04

To Bridge the Semantic Gap...

Recent work, typically

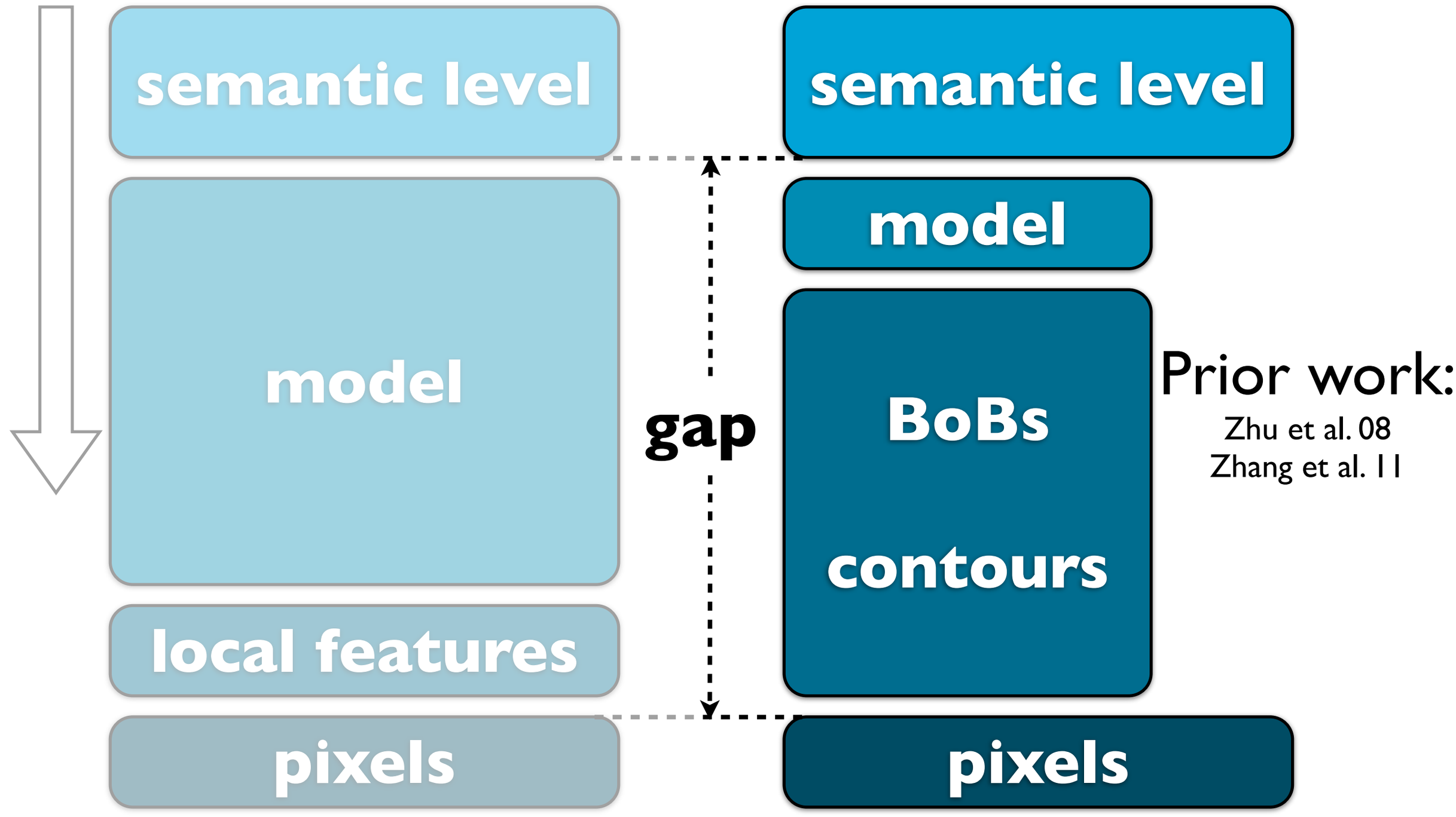
Our approach



To Bridge the Semantic Gap...

Recent work, typically

Our approach



Prior work:

Zhu et al. 08

Zhang et al. 11

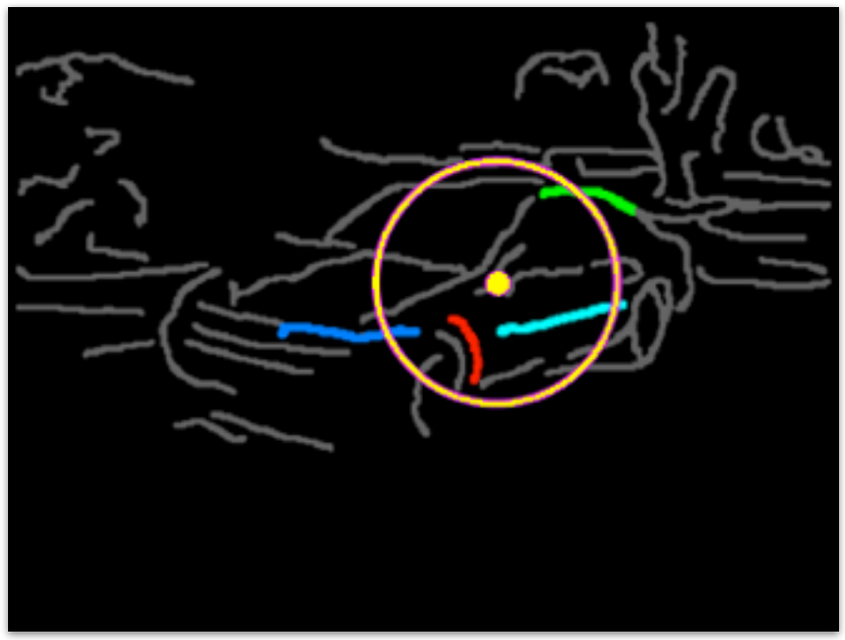
Bags of Boundaries = BoBs

If an object occurs,

it must be in the spotlight of many BoBs

jointly supporting the occurrence hypothesis

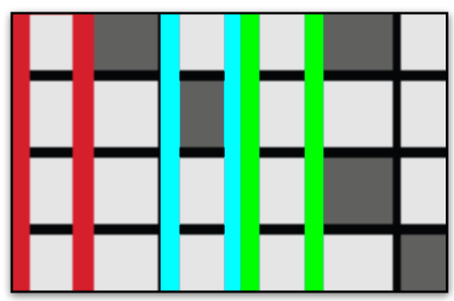
Bags of Boundaries = BoBs

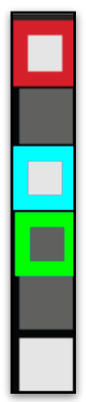


latent indicator of boundaries

shape context

histogram of boundaries = # bins × # contours





contours

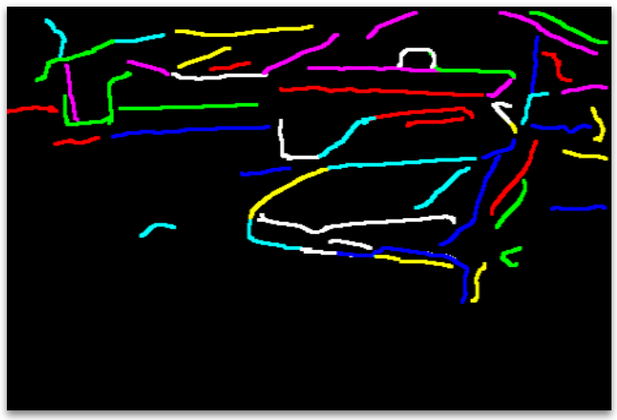
Zhu et al. 08, Zhang et al. 11

Bags of Boundaries vs. Bags-of-Words

BoBs	BoWs
Histogram of **hidden** features that must be inferred	Histogram of **observable** features

Approach

input

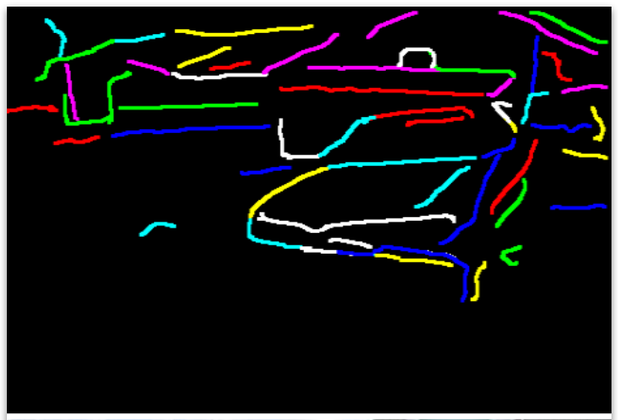


contour extraction

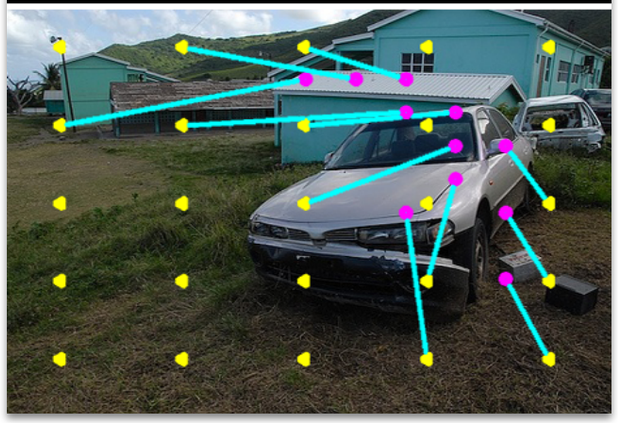
Zhu et al. ICCV07

Approach

input



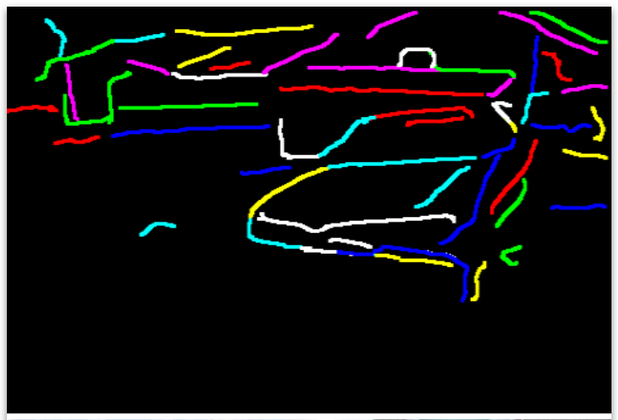
contour extraction



grid of BoBs

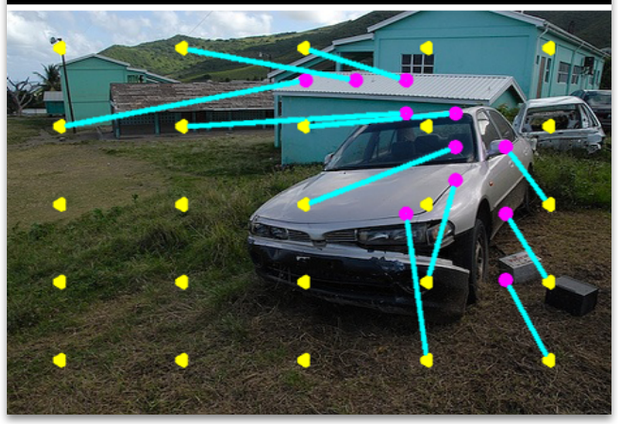
Approach

input



contour extraction

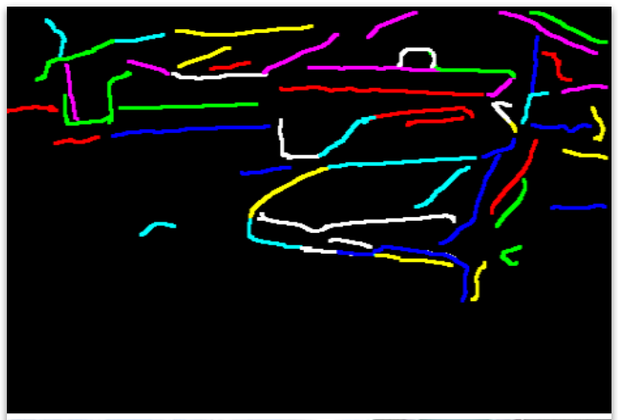
object model



grid of BoBs

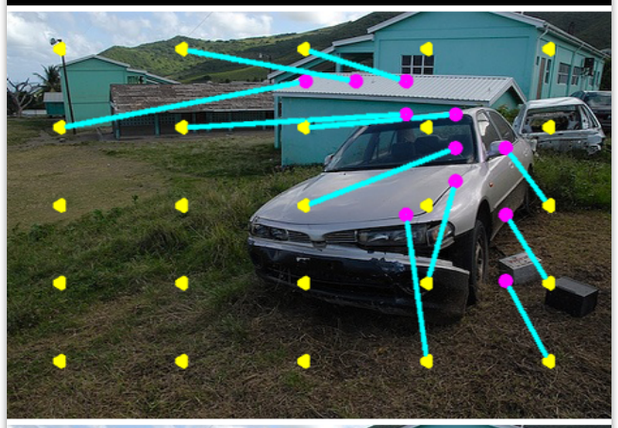
Approach

input

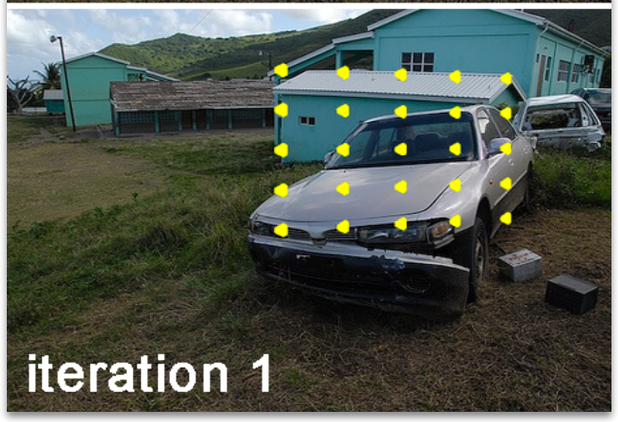


contour extraction

object model



grid of BoBs

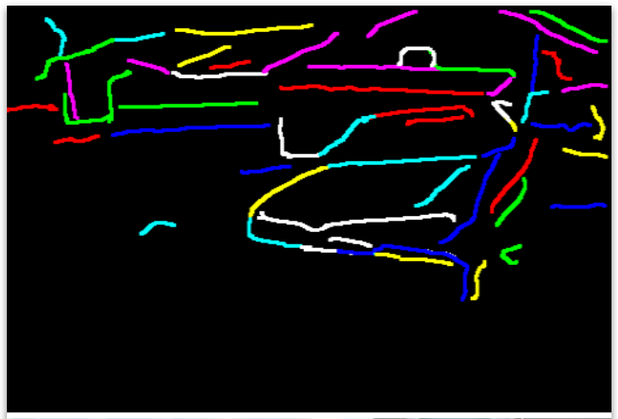


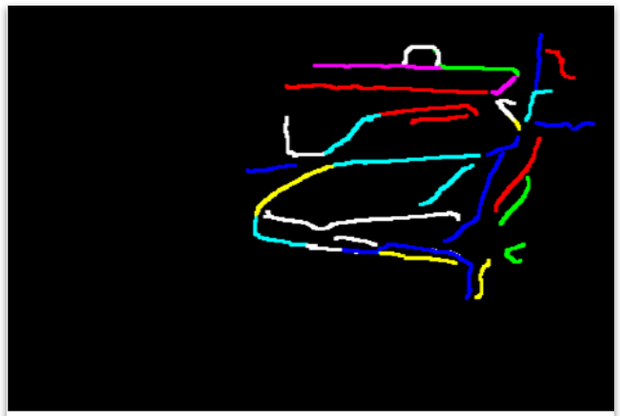
estimate of 3D pose

Approach

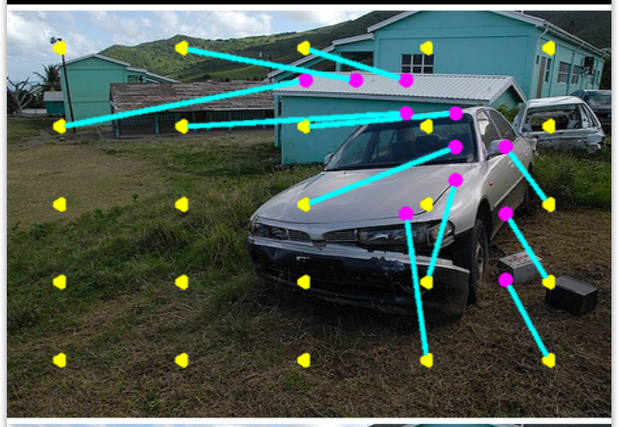
input

object model

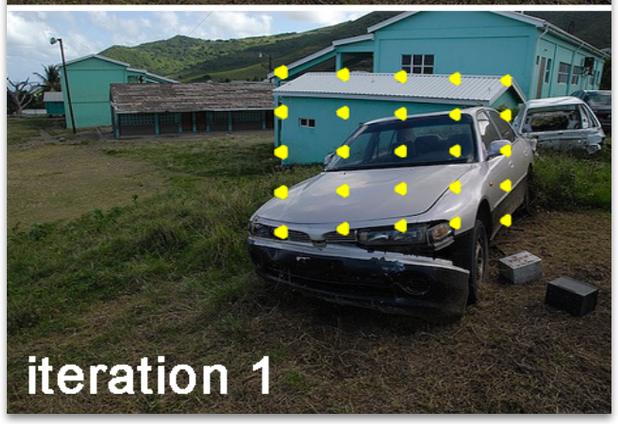




selected boundaries



grid warping

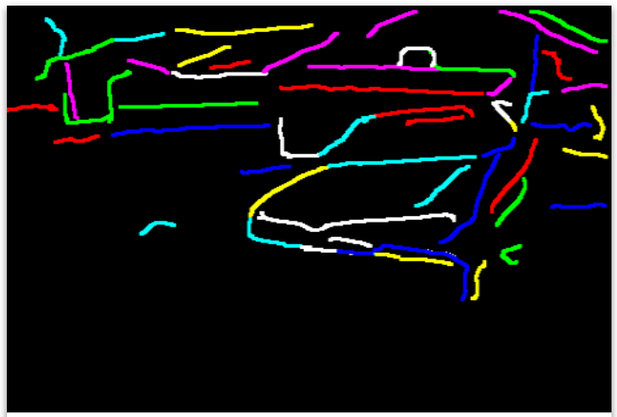


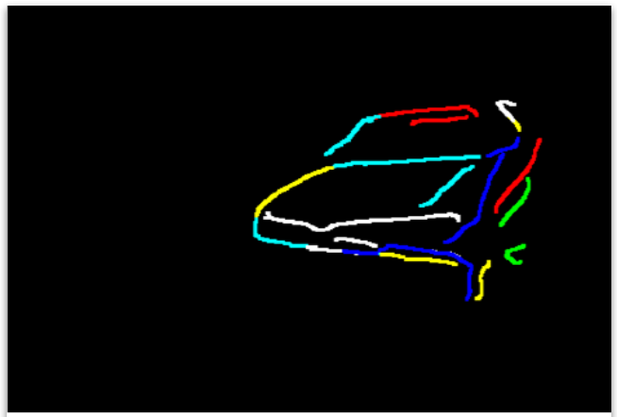
estimate of 3D pose

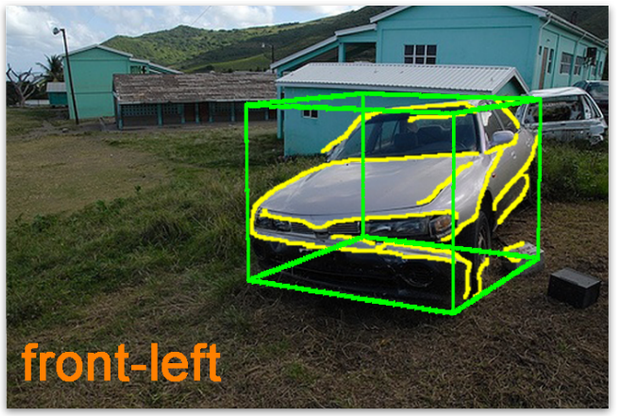
Approach

input

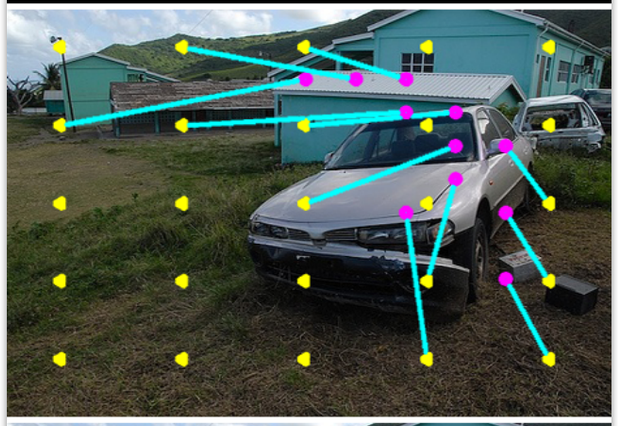
output

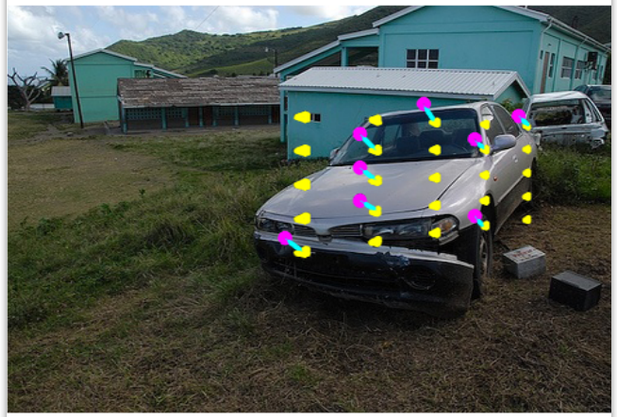


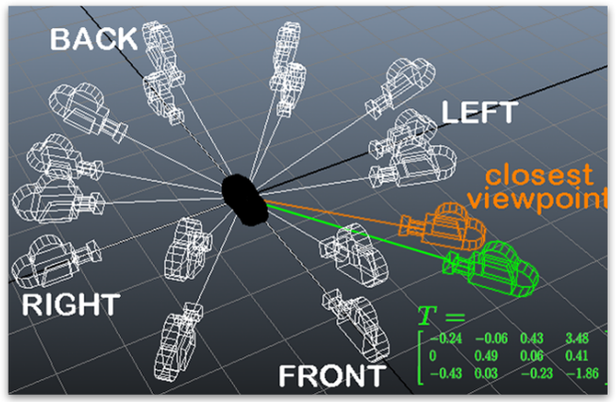


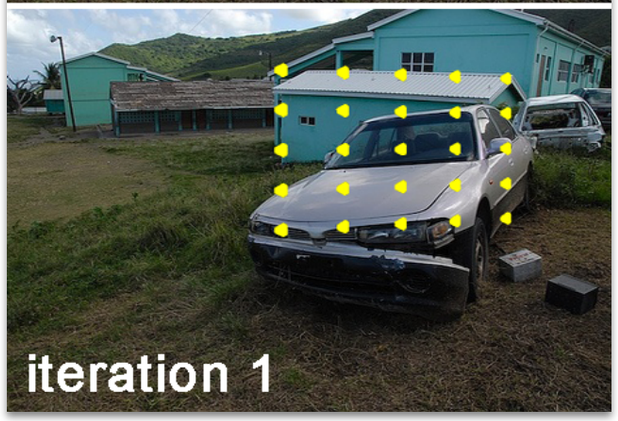


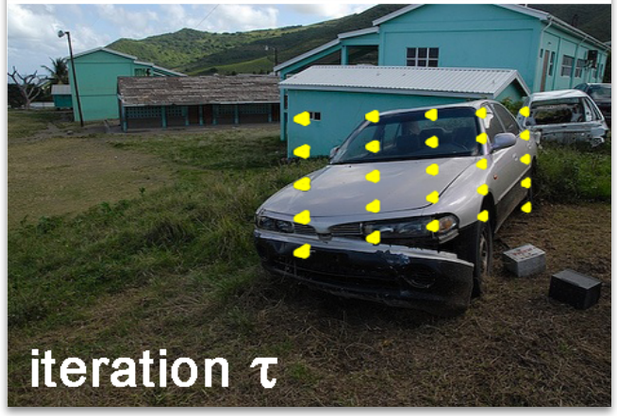
object model











Object Model = Shape Templates

2D probabilistic maps of shape

for a set of viewpoints

Learning

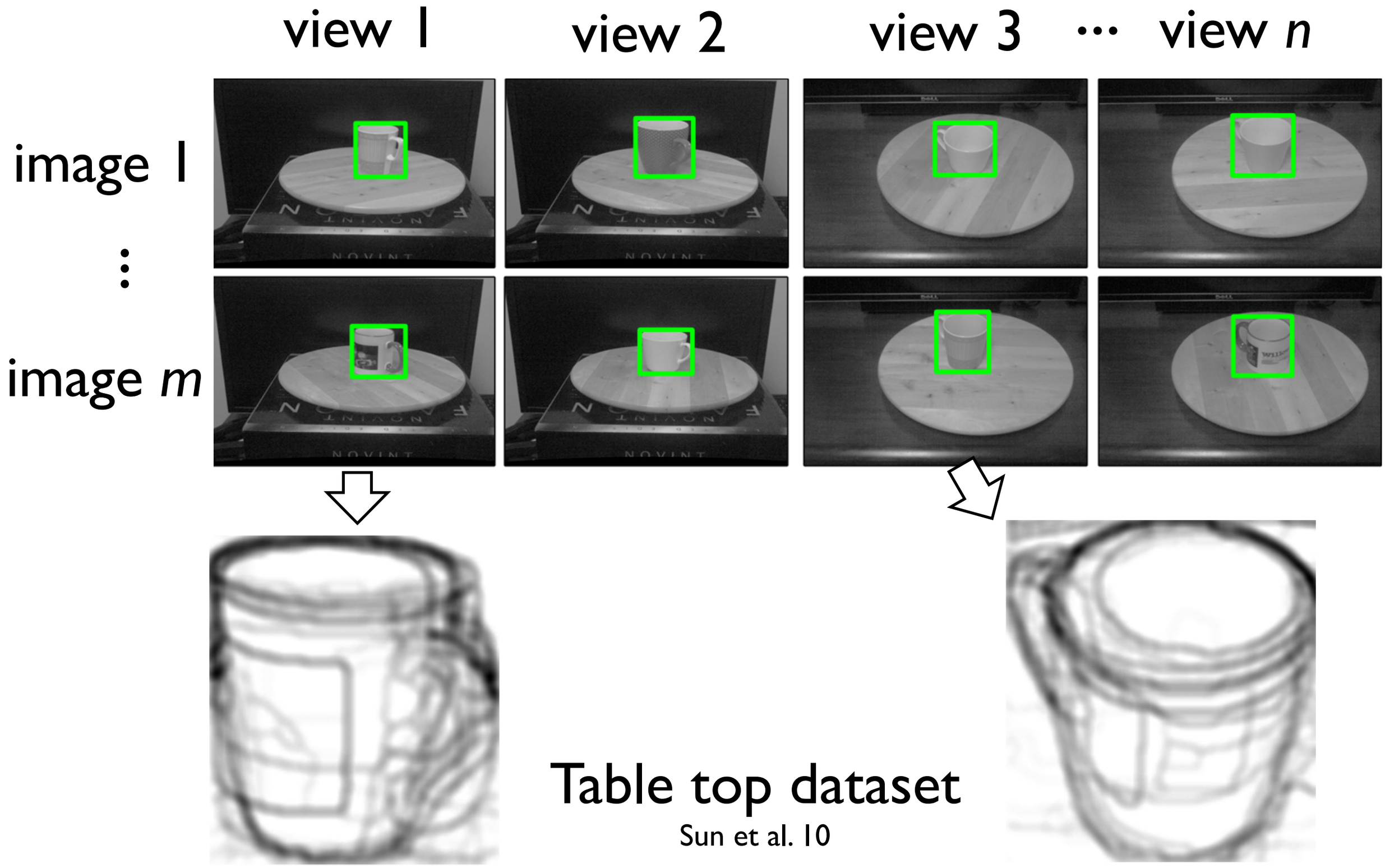
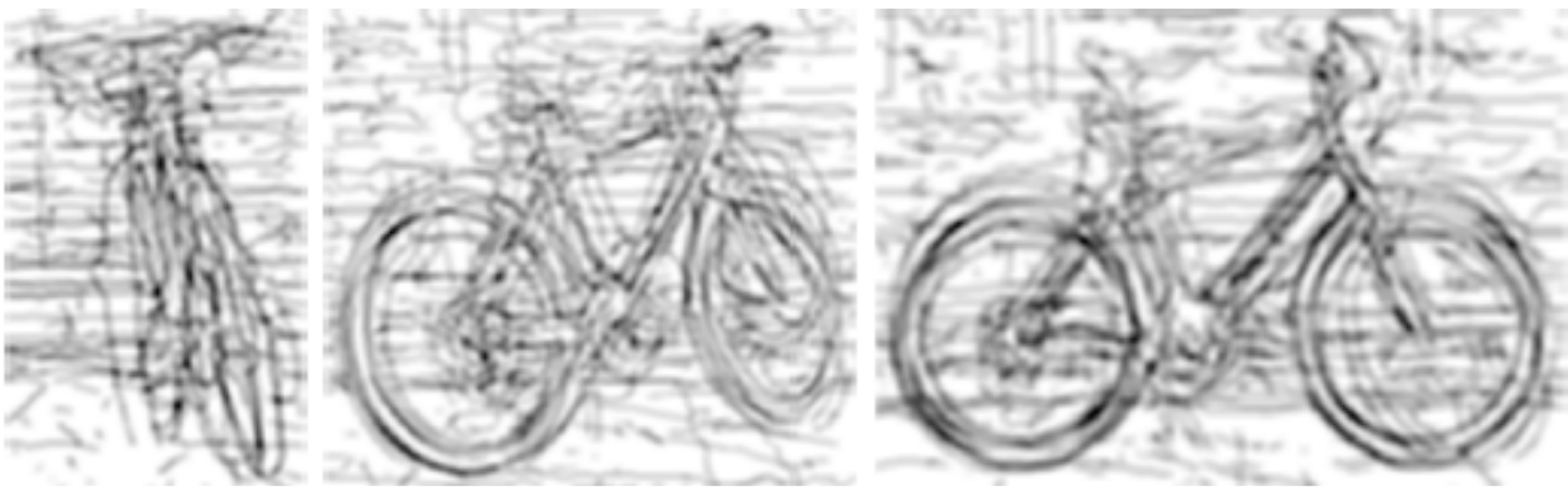


Table top dataset

Sun et al. 10

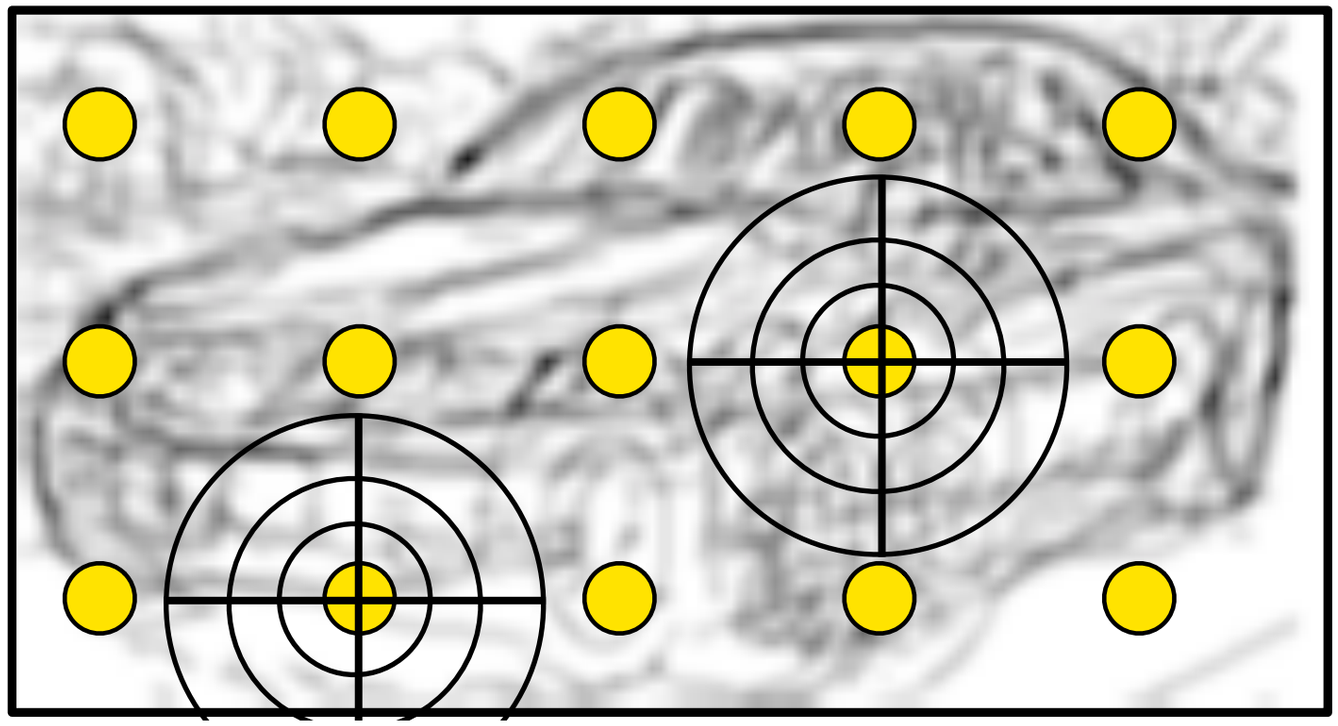
Example Shape Templates



AUTOCAD dataset

Liebelt & Schmid 08-10

Representation of the Shape Template

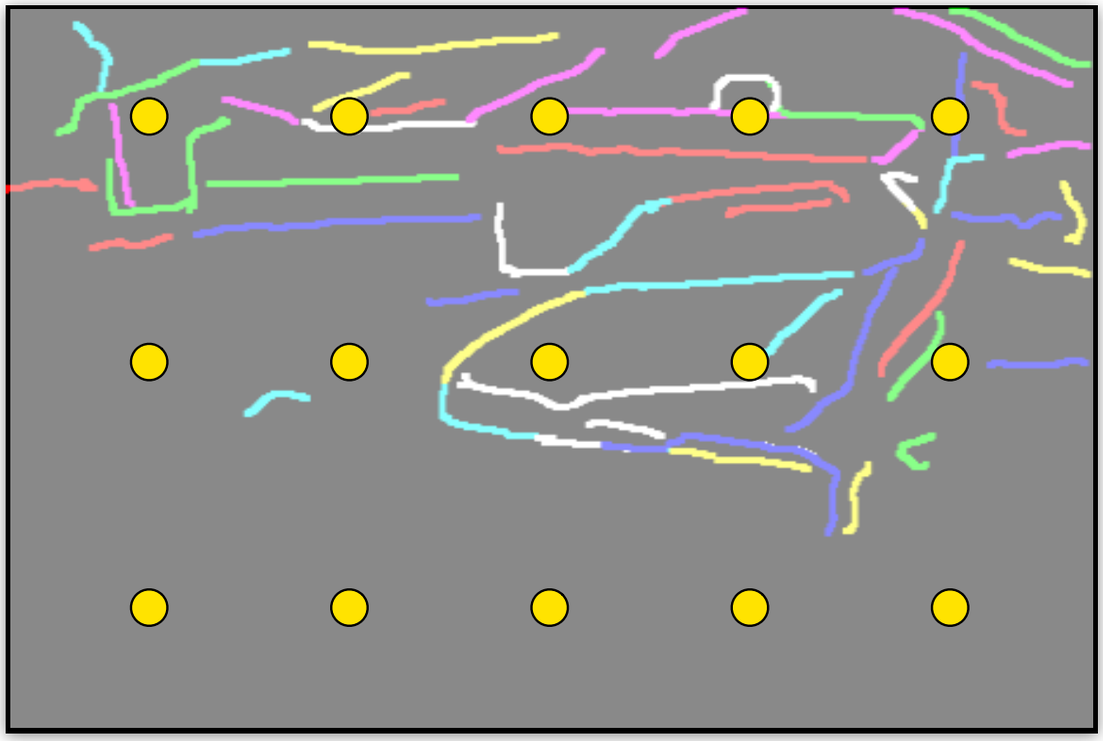


Regular grid of shape-context descriptors

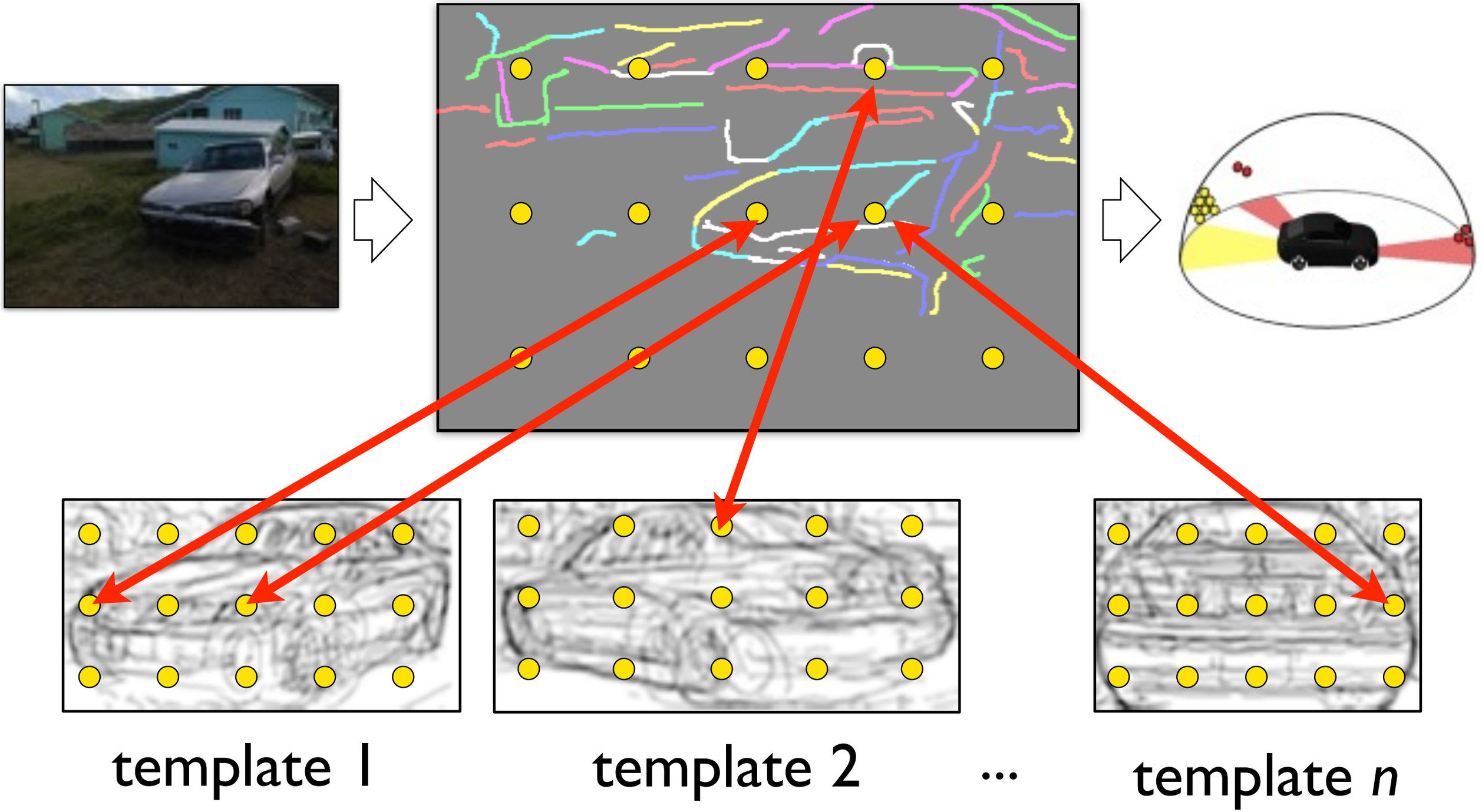
+

Affine projection matrix T

Inference = Matching of BoBs

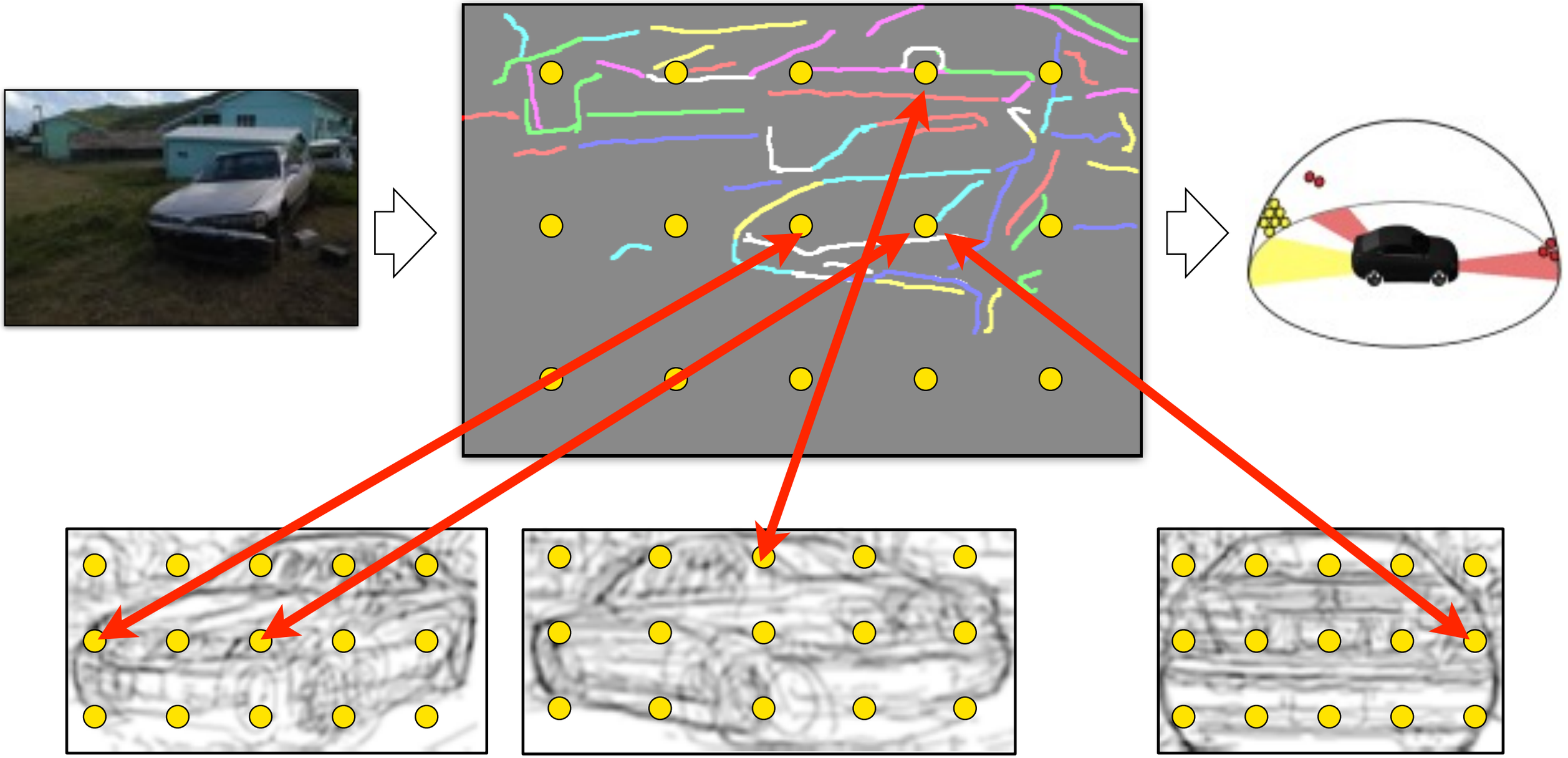


Inference = Matching of BoBs



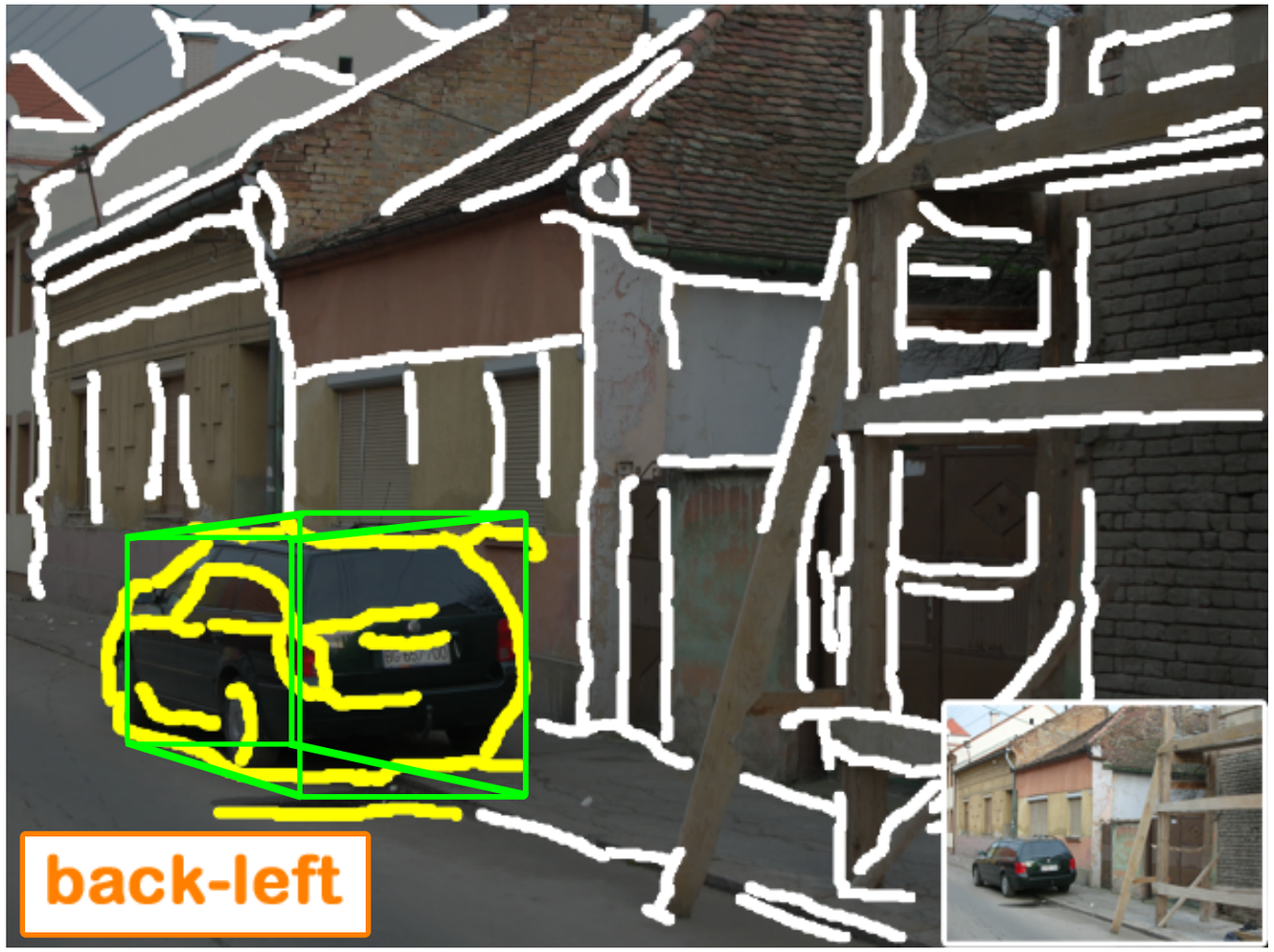
template 1 template 2 ... template *n*

Inference = Matching of BoBs



under an arbitrary affine projection

Example Problem: Object Recognition



Given a set of edges in the image

detect and localize all object instances

and estimate their 3D pose

Payet & Todorovic ICCV11

Matching Formulation

$$\min_{X,F,T} \quad \mathrm{tr}\left\{C^T(X)F\right\} + \alpha||TQF^T - P|| + \beta||(TQF^T - P) - (TQF^T - P)W^T||$$

Matching Formulation

$$\min_{X,F,T} \quad \mathrm{tr}\left\{C^T(X)F\right\} + \alpha||TQF^T - P|| + \beta||(TQF^T - P) - (TQF^T - P)W^T||$$

$$\text{s.t. } X \in [0,1]^N$$

Matching Formulation

$$\min_{X,F,T} \operatorname{tr}\{C^T(X)F\} + \alpha||TQF^T - P||$$

$$+\beta||(TQF^T - P) - (TQF^T - P)W^T||$$

$$\text{s.t. } F \geq 0; F^T\mathbf{1}_N = \mathbf{1}_M; F\mathbf{1}_M \leq \mathbf{1}_N$$

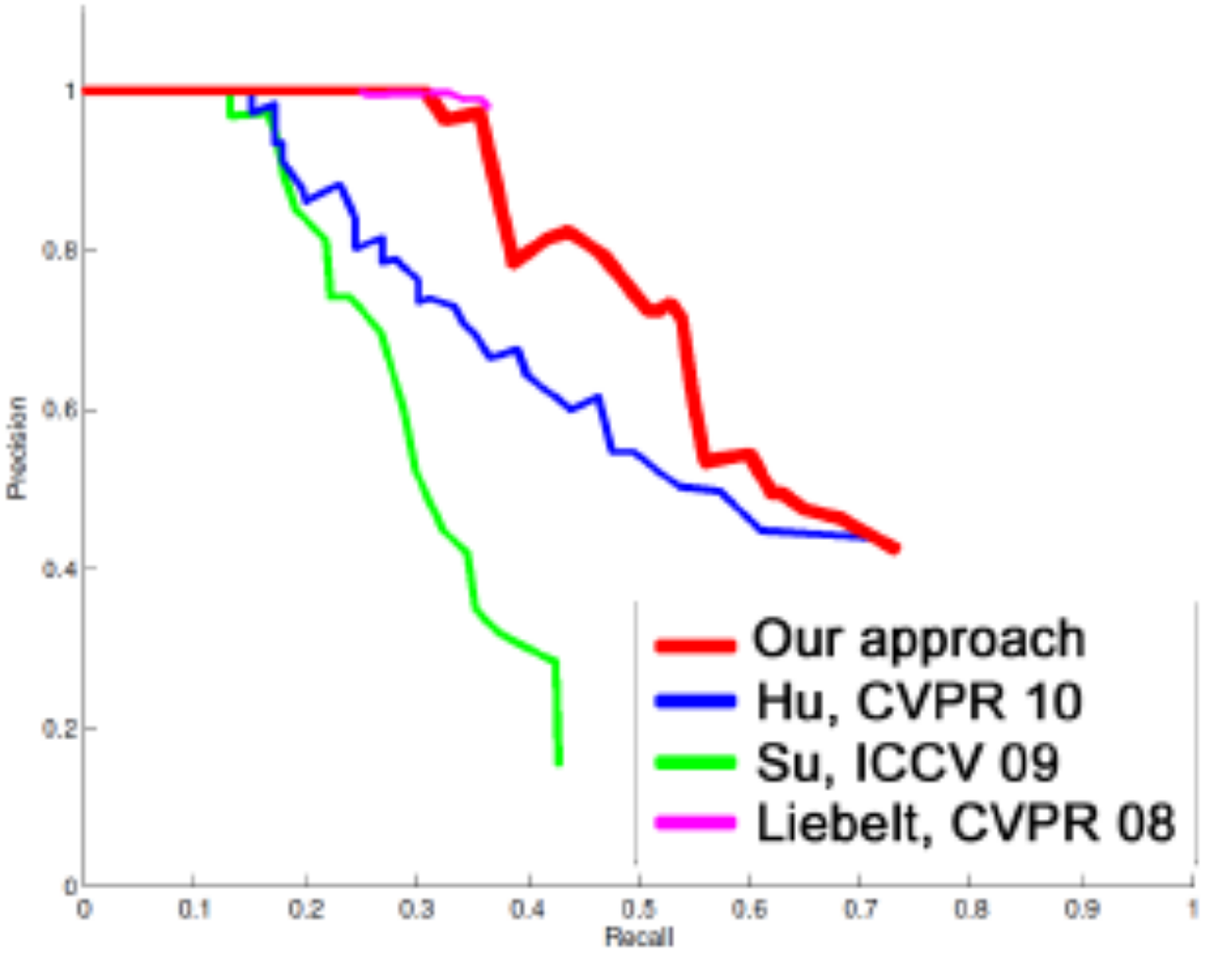
Matching Formulation

$$\min_{X,F,T} \operatorname{tr}\{C^T(X)F\} + \alpha||TQF^T - P|| +$$

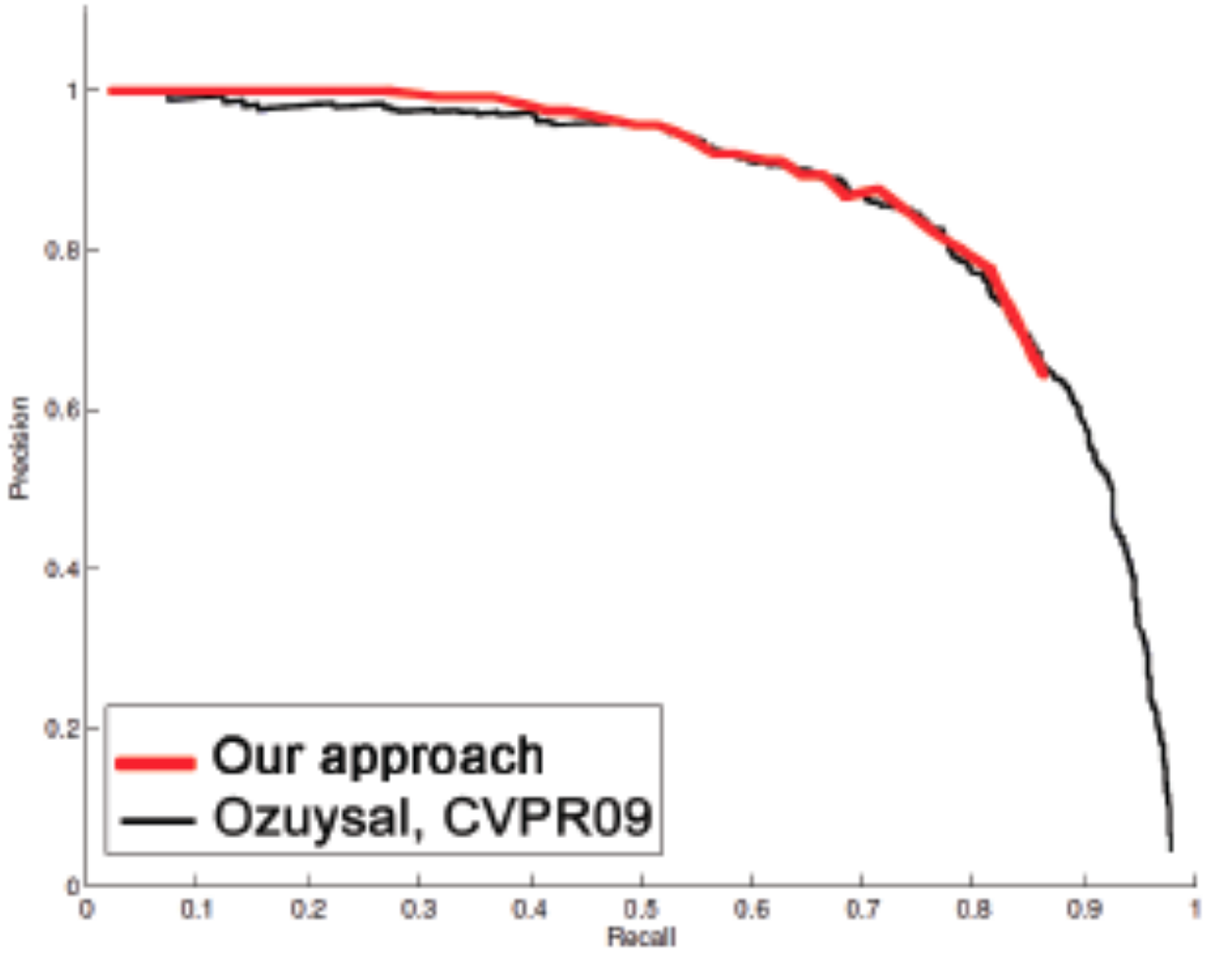
$$+\beta||(TQF^T - P) - (TQF^T - P)W^T||$$

$$\text{s.t. } F \geq 0; F^T\mathbf{1}_N = \mathbf{1}_M; F\mathbf{1}_M \leq \mathbf{1}_N$$

Results: Object Detection

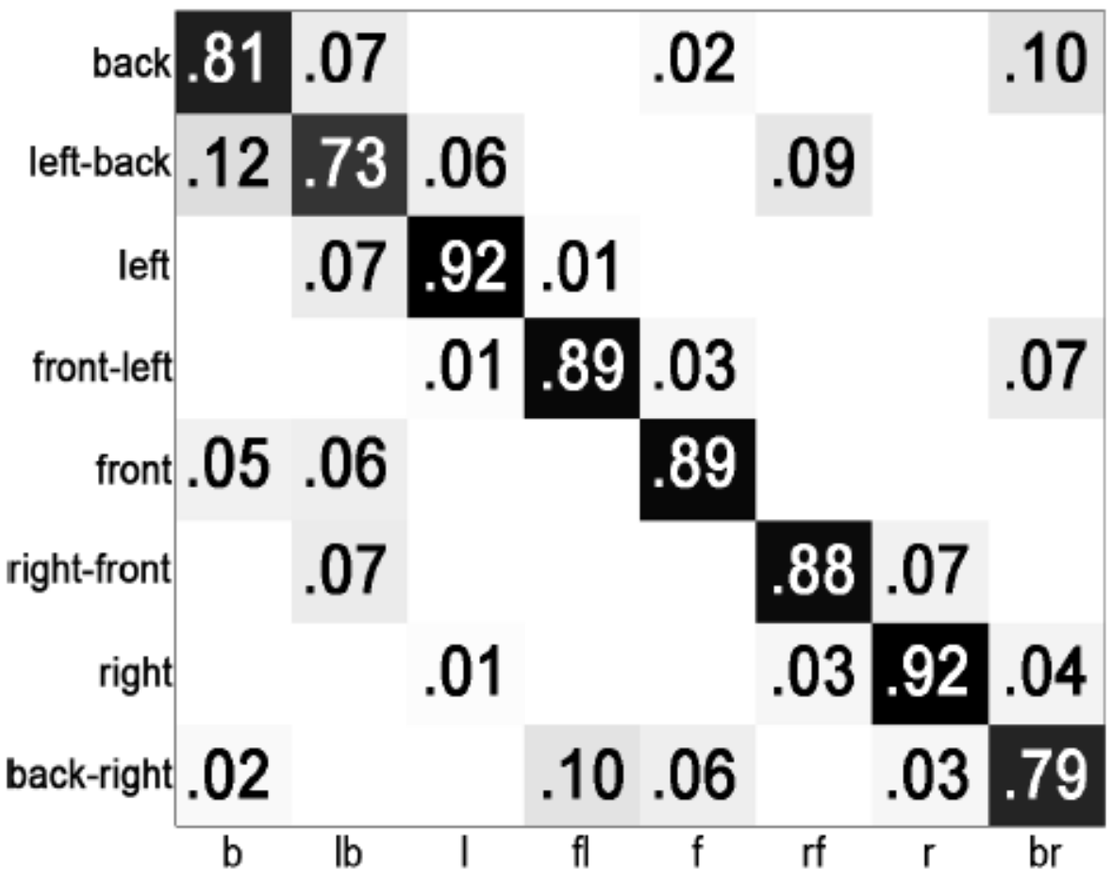


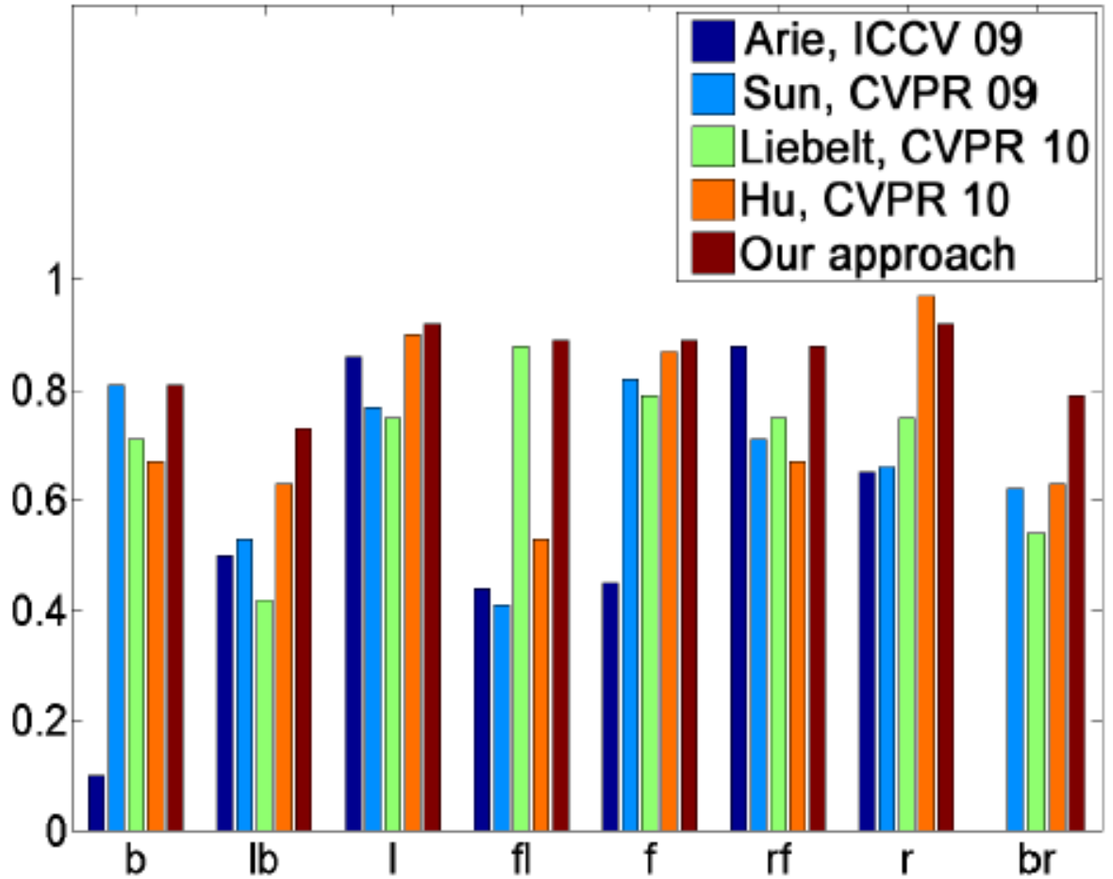
PASCAL VOC 2006
car dataset



Car show dataset

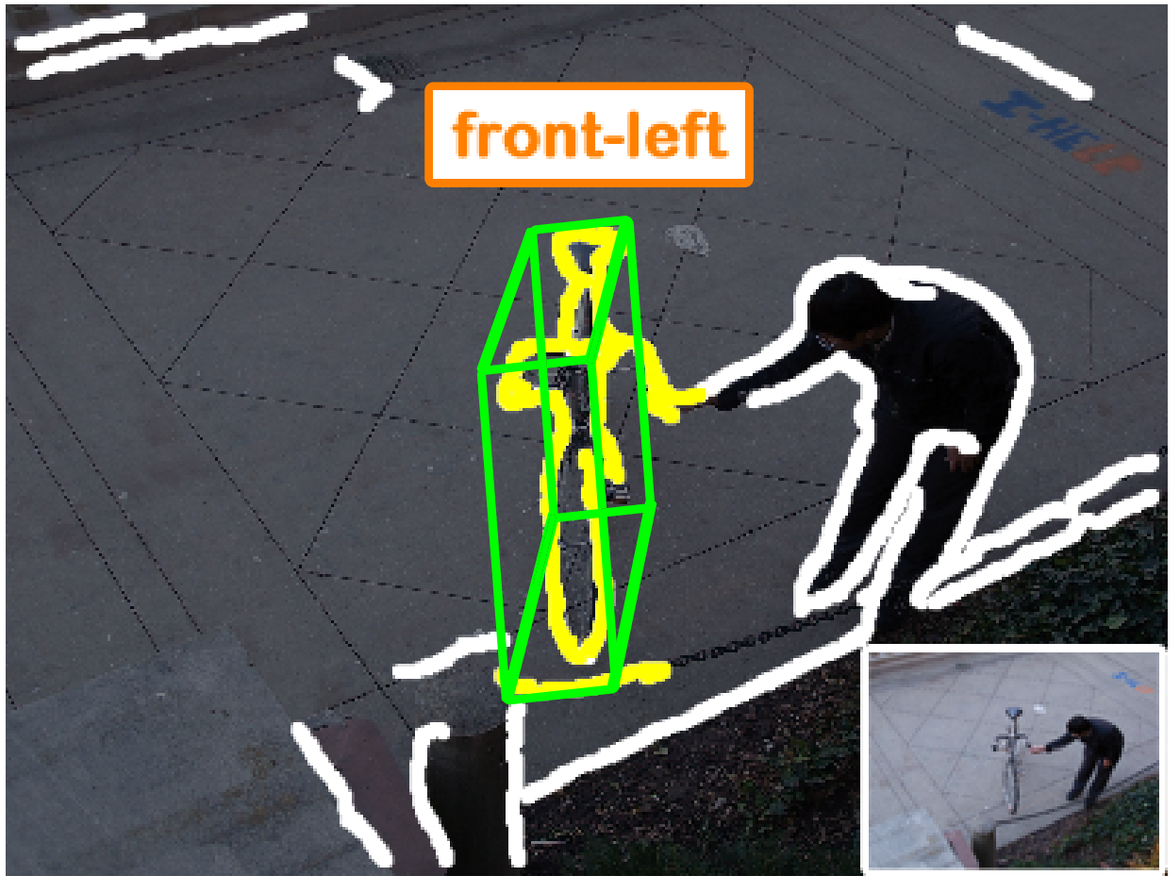
Results: Viewpoint Classification

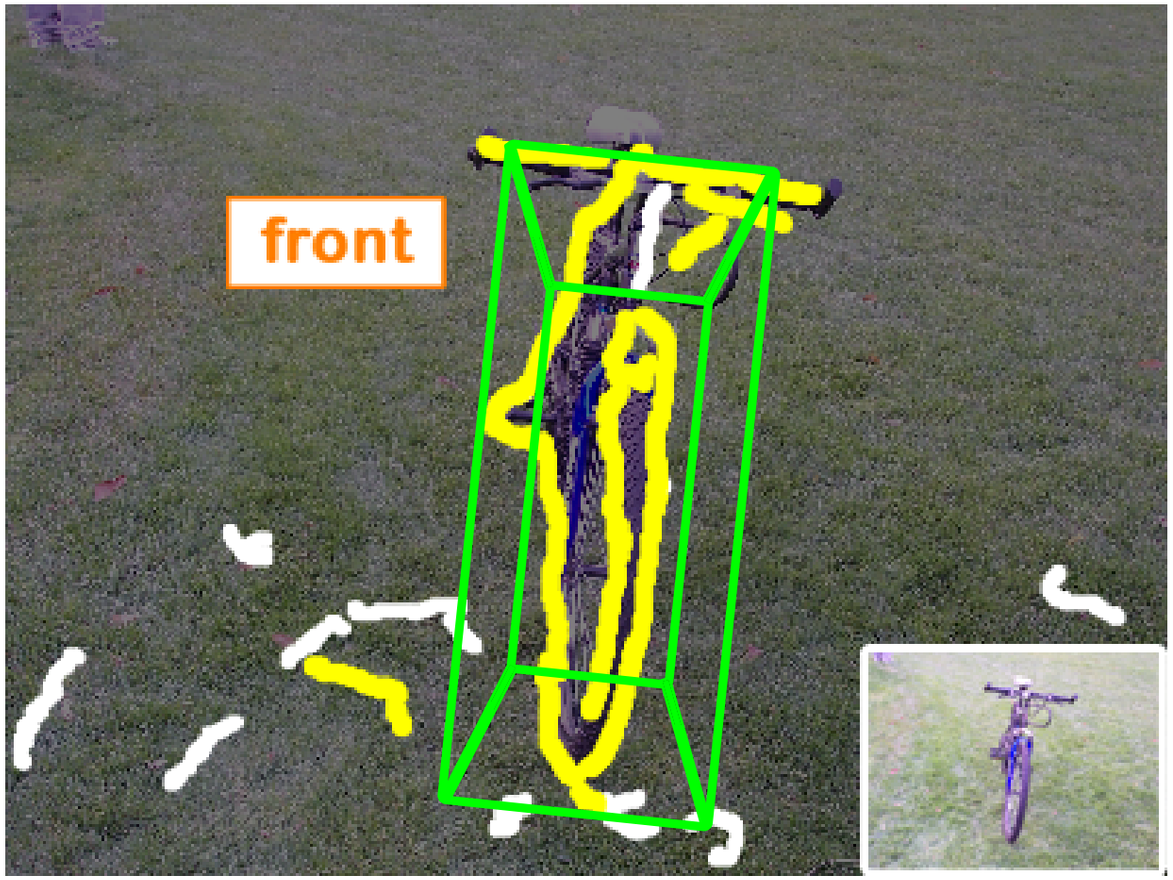




3D Object dataset: Cars

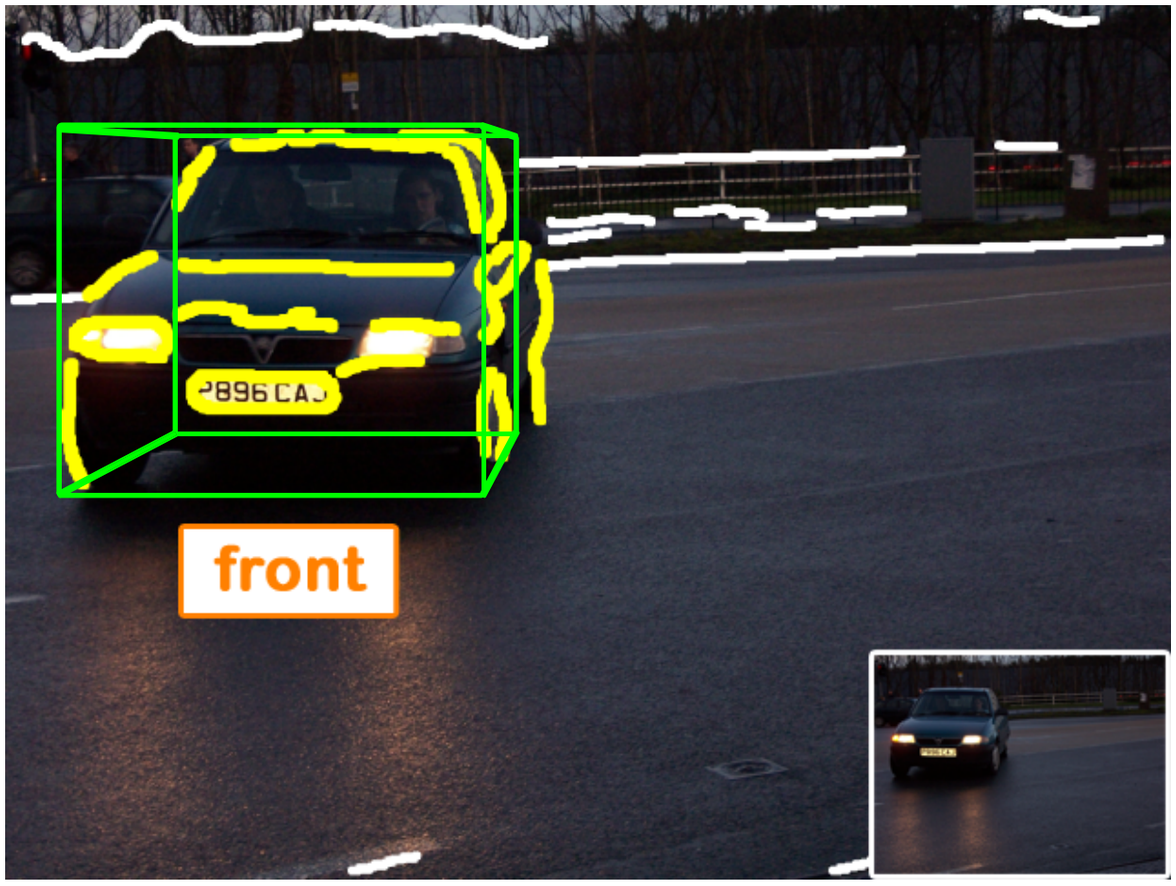
Results: 3D Pose Estimation

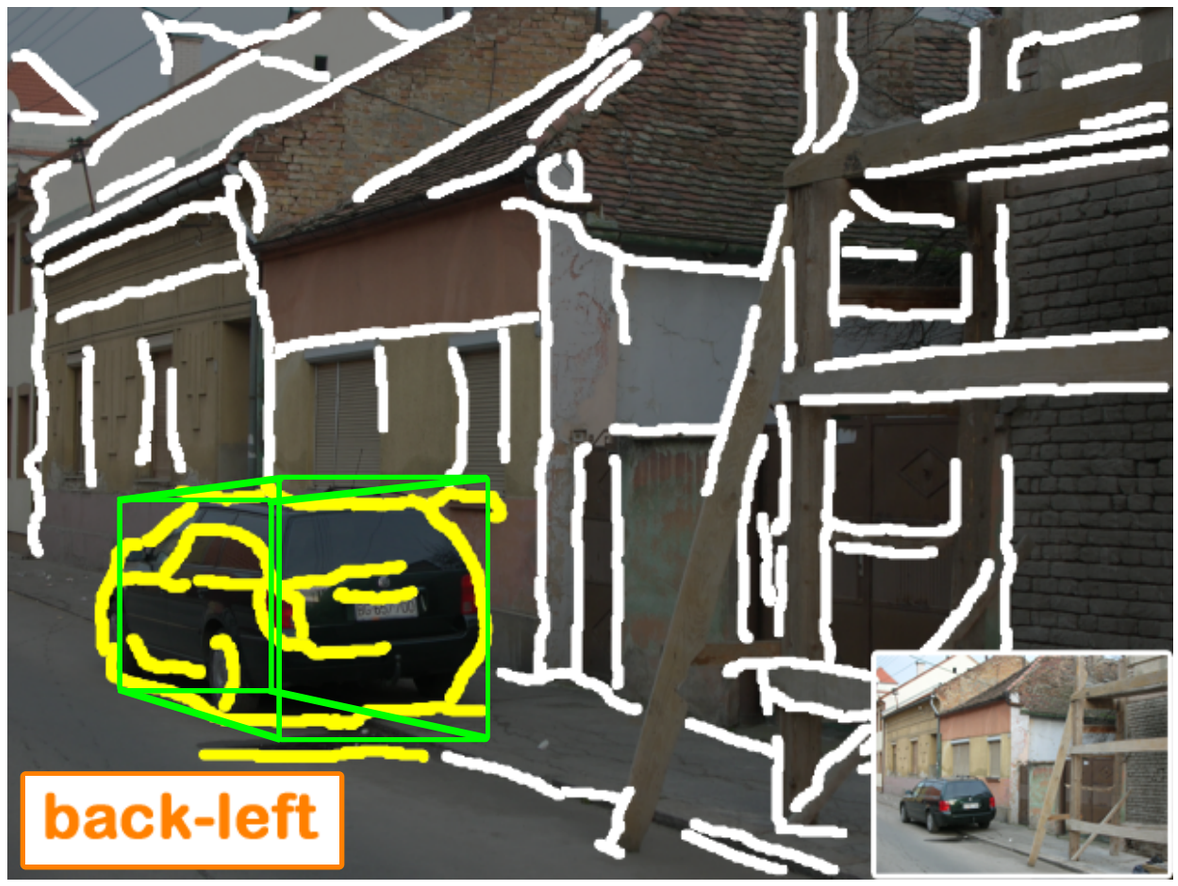




Correct detection, localization, and pose estimation

Results: 3D Pose Estimation





Correct detection, localization, and pose estimation

Conclusion

- Recent work:
  - Pre-selected local features
  - Sophisticated object models and algorithms
- Our approach:
  - Mid-level features allow for:
    - Abstracting low-level features
    - Synergistic bottom-up/top-down interaction
  - Simple models and algorithms