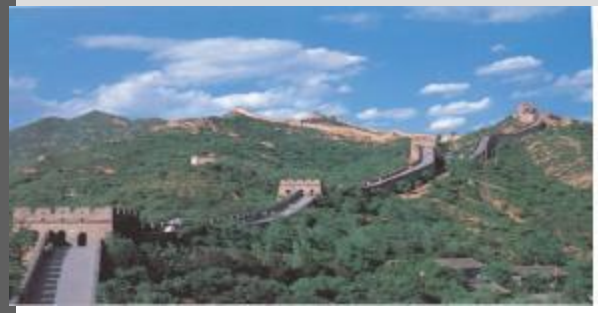
Methods for Representing and Recognizing 3D objects

part 2



1st Sino-USA Summer School in Vision, Learning, and Pattern Recognition

VLPR 2009 • July 20-27, 2009 • Peking University, Beijing, China

Silvio Savarese

University of Michigan at Ann Arbor

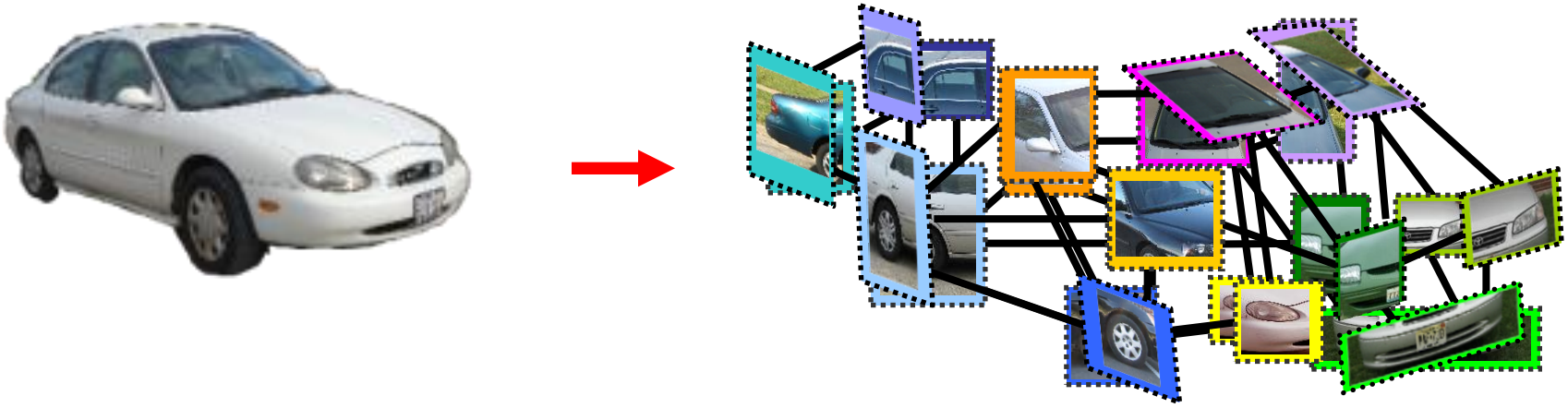
Part- based Multi-view models

Savarese, Fei-Fei, ICCV 07

Savarese, Fei-Fei, ECCV 08

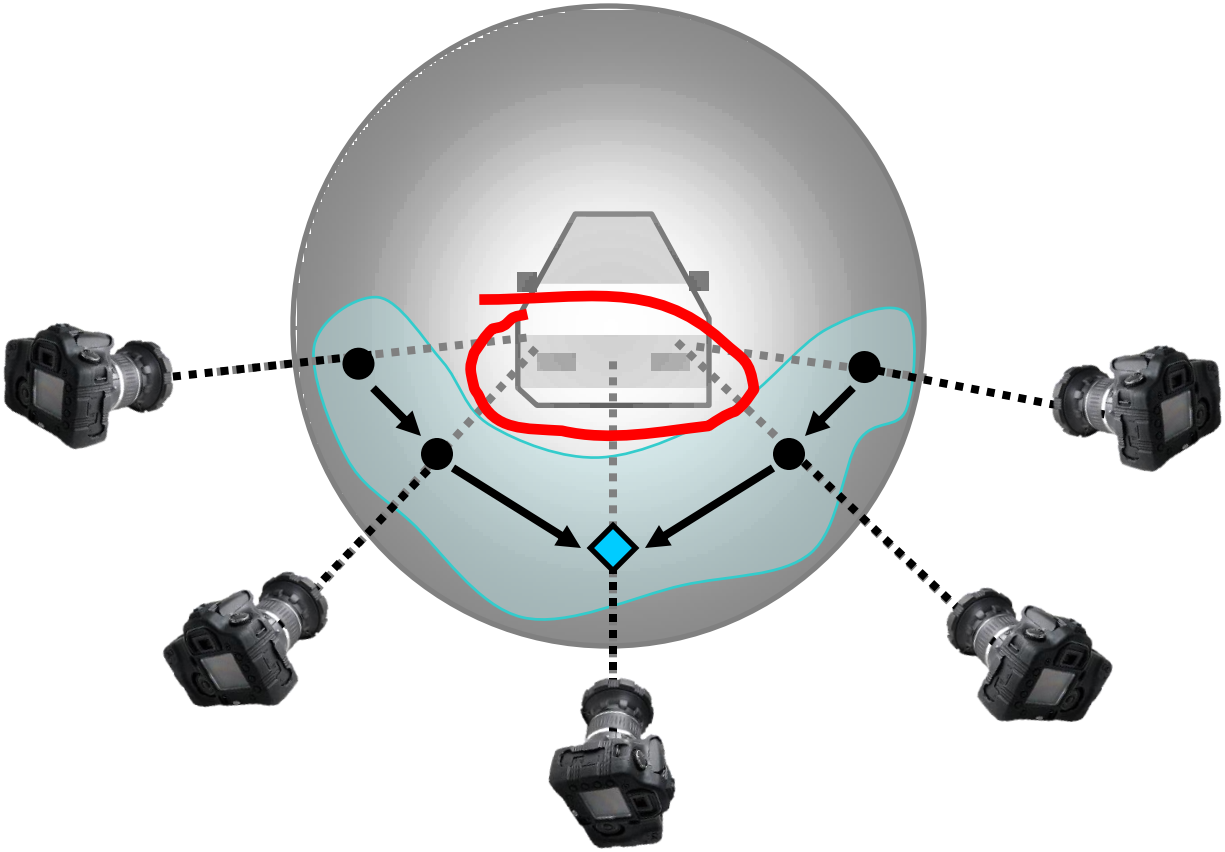
Sun, Su, Savarese, Fei-Fei, CVPR 09

Su, Sun, Fei-Fei, Savarese, ICCV 09



- Canonical parts captures diagnostic appearance information
- 2d ½ structure linking parts via weak geometry

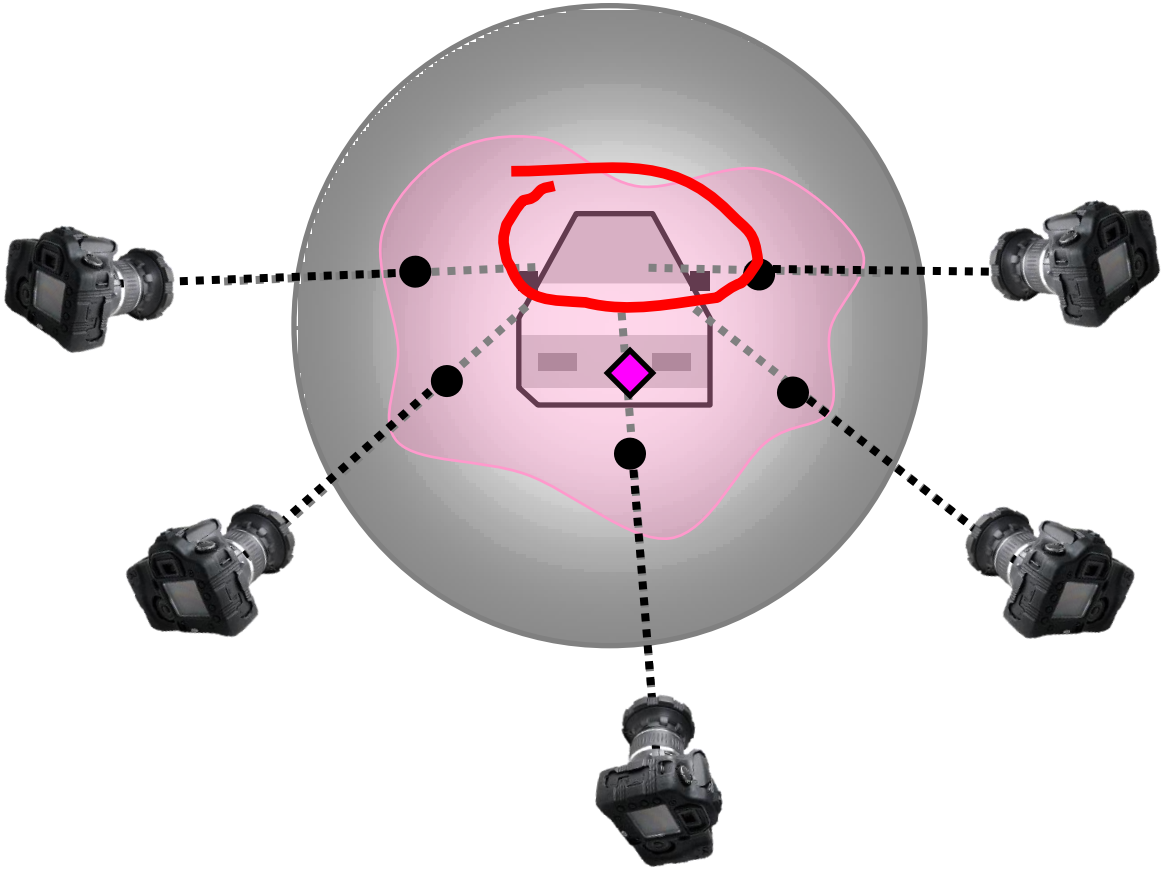
Canonical parts



- If physical part is planar, canonical part is stable point on the manifold
- Canonical part can be computed from connected component of parts

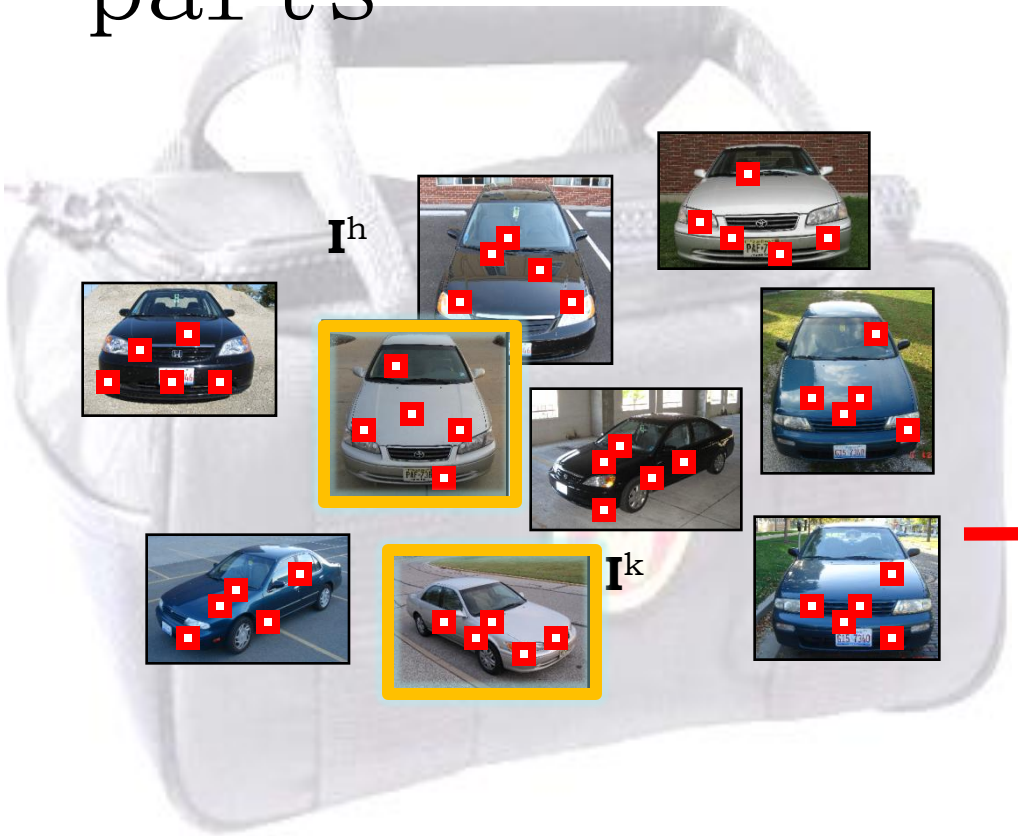
connected
component of parts

Canonical parts



connected
component of parts

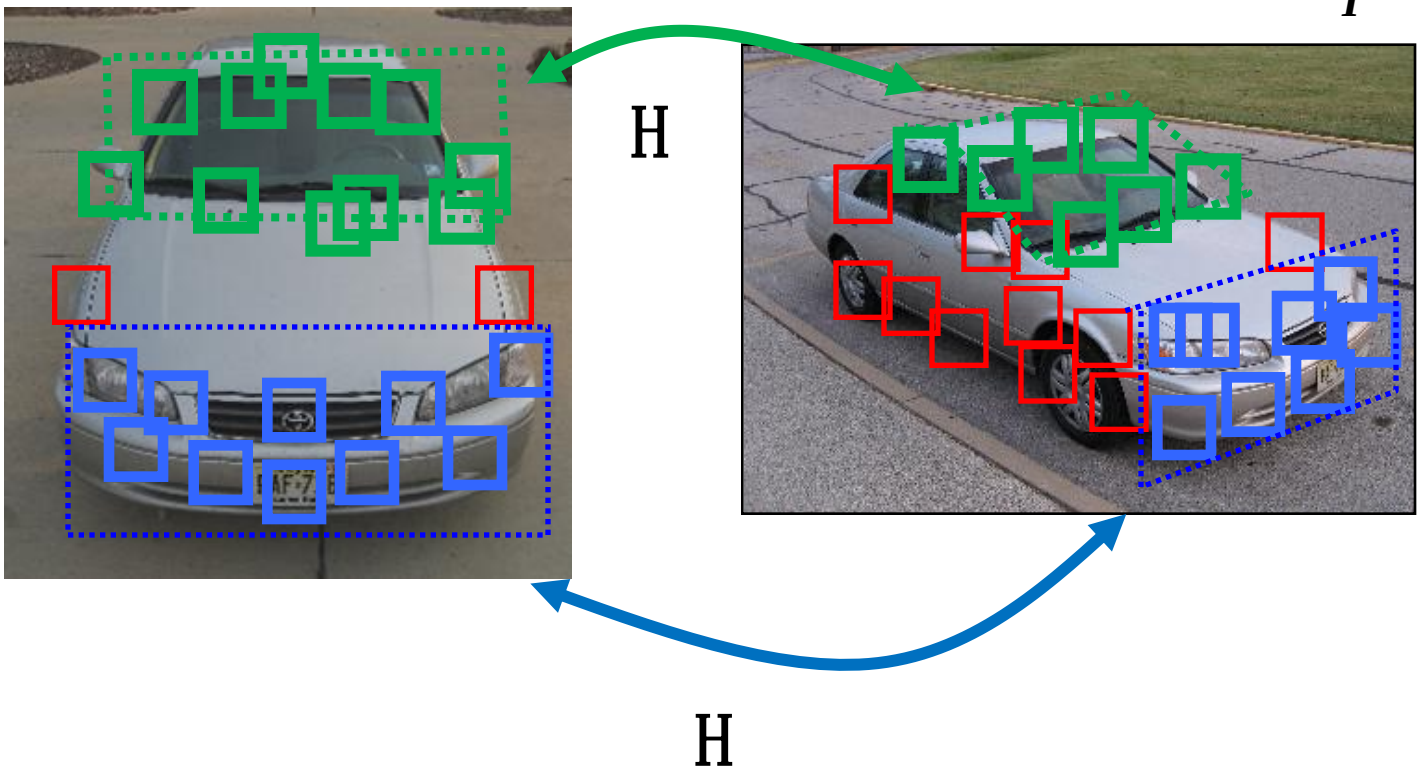
Connected components of parts



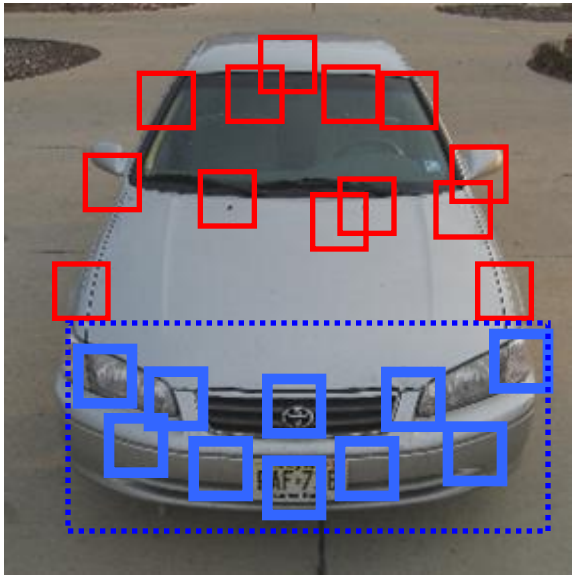
Unlabeled mix of images:

- category labels
- no pose labels;
- images of same instance from multiple views

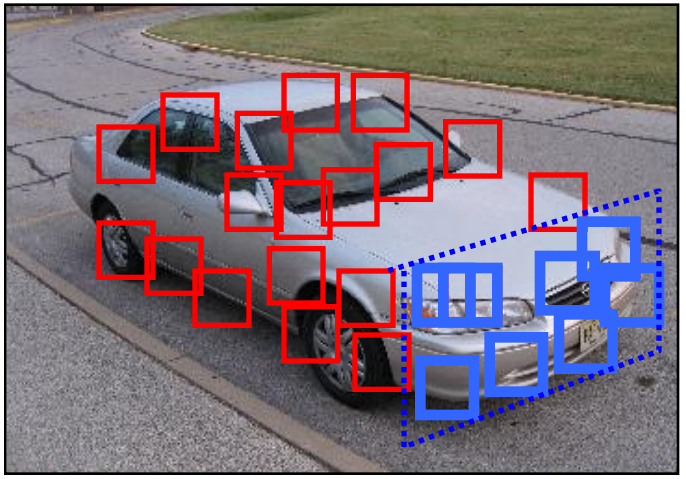
$I^h = [x_1, x_2, \cdots, x_M]$



$I^h = [x_1, x_2, \cdots, x_M]$



I^k



$x_{2,1}, x_{2,2}, \cdots, x_{2,n}$

$$\pi : I^h \rightarrow \left\{ P_1^h, P_2^h, \cdots P_N^h, O^h \right\}$$

$$\tau : I^k \rightarrow \left\{ P_1^k, P_2^k, \cdots P_N^k, O^k \right\}$$

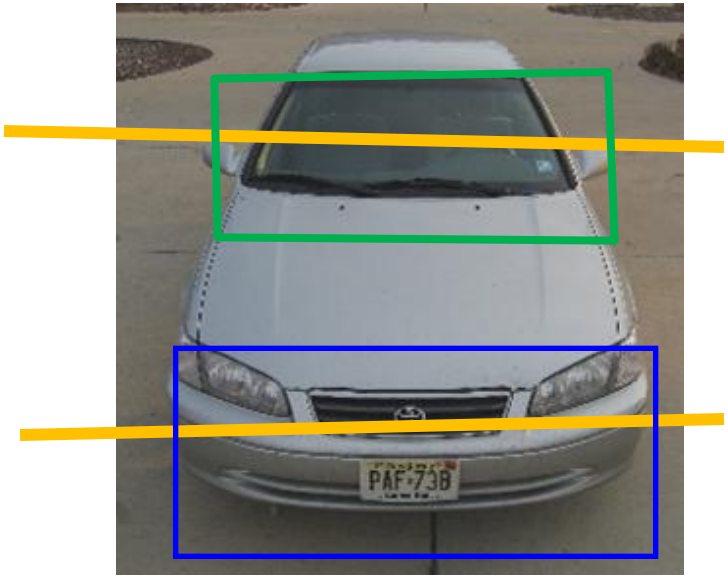
$$\min \left| O^h \cup O^k \right| + c\ N$$

such that:

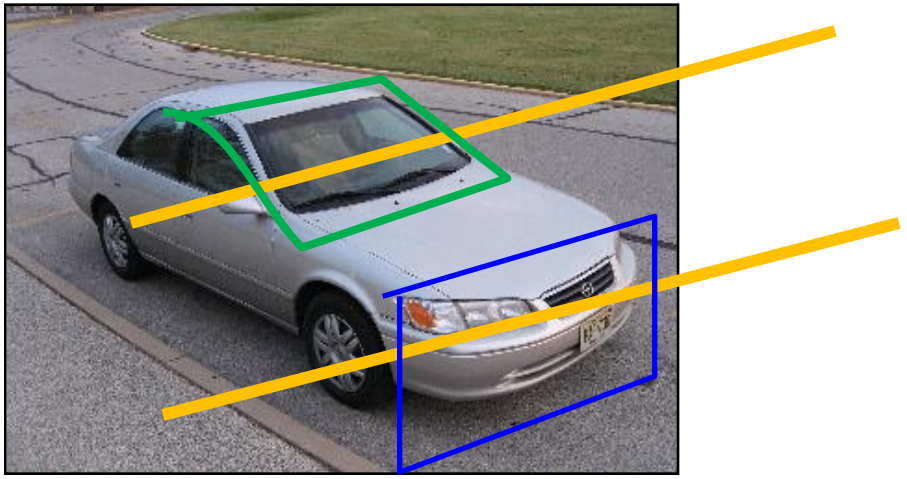
$$f(P_i^h, P_j^k, \beta_{i,j}) < \delta \qquad \forall \text{ i, j} = 1 \cdots \text{N}$$

$$f(P_i^h, P_j^k, \beta_{i,j}) = \left\| P_i^h - H_{i,j}\ P_j^k \right\|$$

$I^h = [x_1, x_2, \cdots, x_M]$



I^k



$$\pi : I^h \rightarrow \left\{ P_1^h, \; P_2^h, \cdots \; P_N^h, \; O^h \right\}$$

$$\tau : I^k \rightarrow \left\{ P_1^k, \; P_2^k, \cdots \; P_N^k, \; O^k \right\}$$

$$\min \left| O^h \cup O^k \right| + c \; N$$

such that:

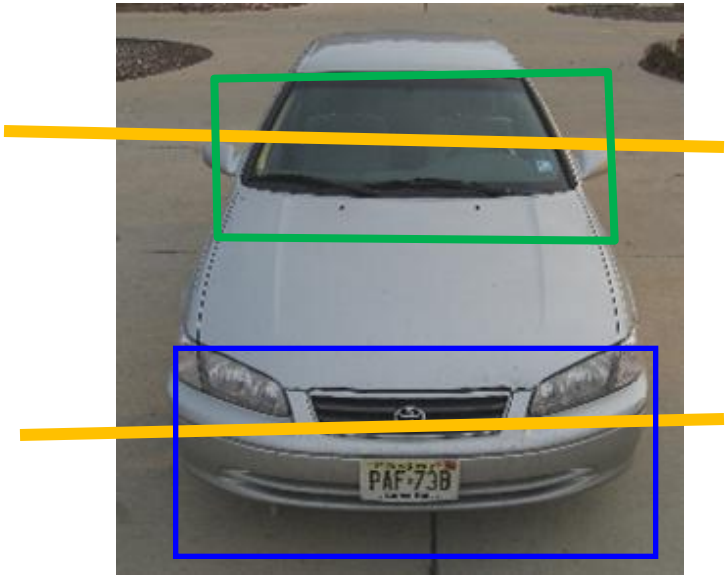
$$f(P_i^h, P_j^k, \beta_{i,j}) < \delta \qquad \forall \, i, \, j = 1 \cdots N$$

$$f(P_i^h, P_j^k, \beta_{i,j}) = \left\| P_i^h - H_{i,j} \; P_j^k \right\|$$

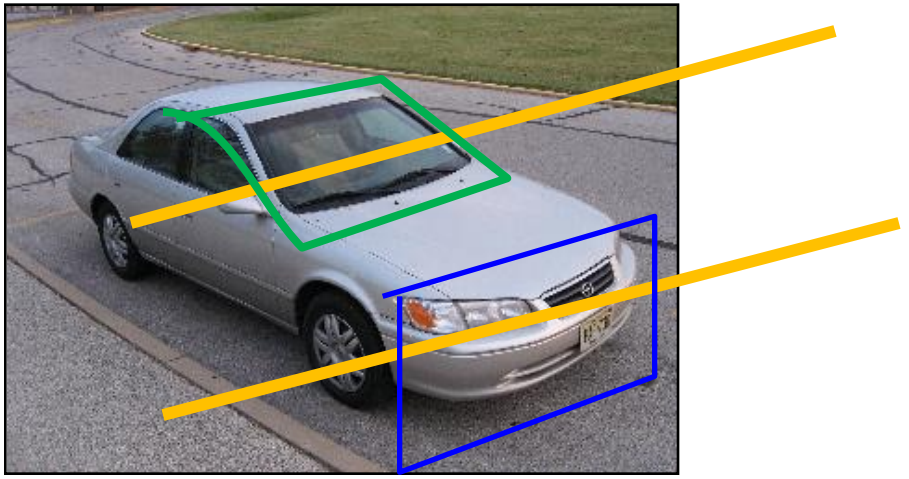
$$f(P^h, P^k, \gamma) < \delta$$

$$f(P^h, P^k, \gamma) = \left\| \left[P_1^h, \cdots P_N^h \right] F \left[P_1^k, \cdots P_N^k \right]^T \right\|$$

$I^h = [x_1, x_2, \cdots, x_M]$



I^k



$$\pi : I^h \rightarrow \left\{ P_1^h, \; P_2^h, \cdots \; P_N^h, \; O^h \right\}$$

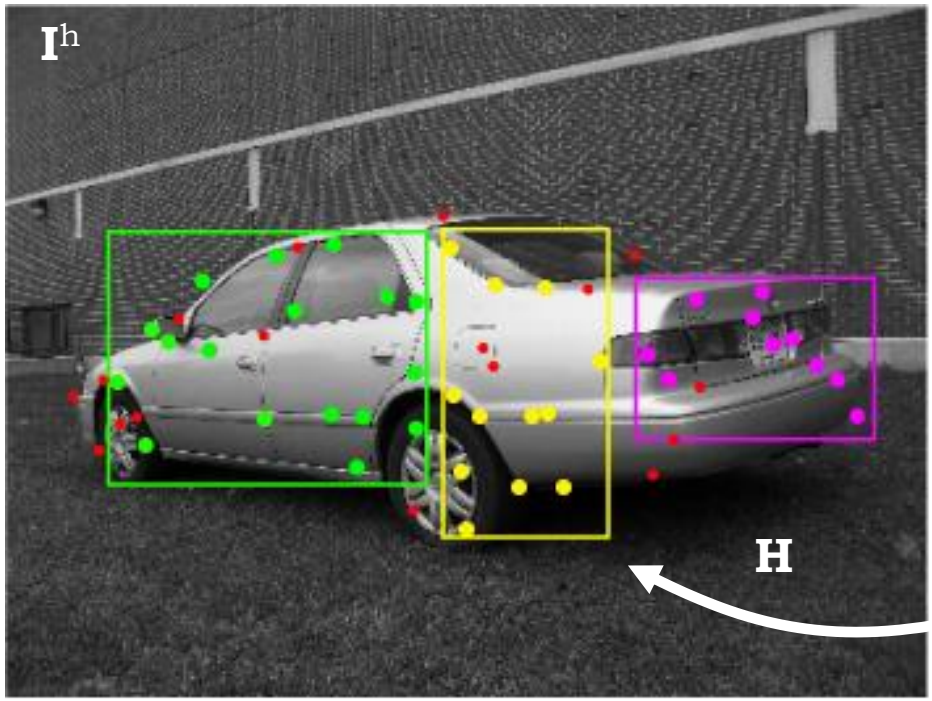
$$\tau : I^k \rightarrow \left\{ P_1^k, \; P_2^k, \cdots \; P_N^k, \; O^k \right\}$$

$$\min \; \left| O^h \cup O^k \right| + c \; N$$

GOAL:

- discover partition while fitting multiple homographies
- fit global constraint
- minimize outlier set

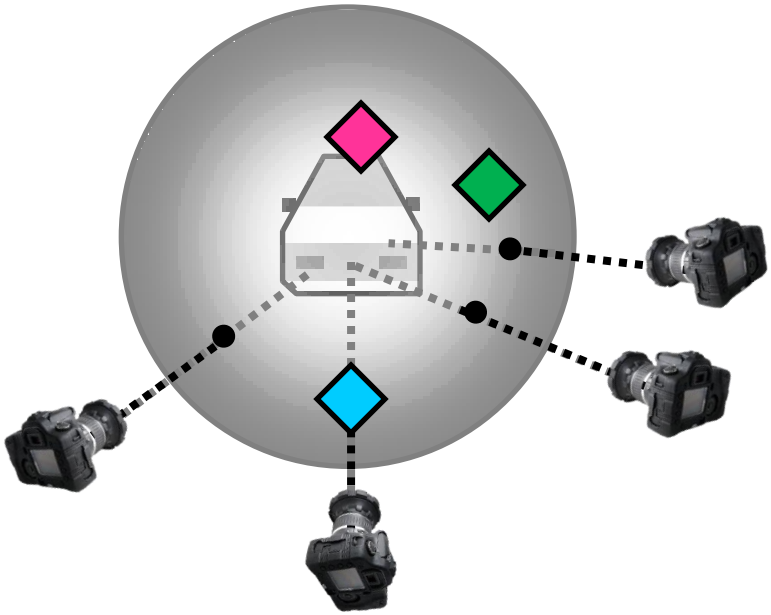
- Use sequential RANSAC or RANSAC & J-linkage [toldo, fusiello eccv 08]

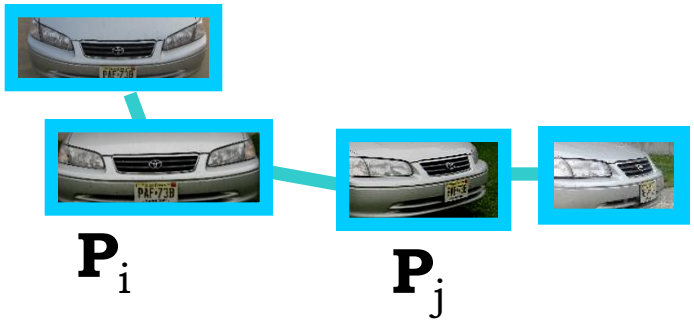


$$\pi : \boldsymbol{I}^{\boldsymbol{h}} \rightarrow \left\{ \boldsymbol{P}_1^{\boldsymbol{h}}, \boldsymbol{P}_2^{\boldsymbol{h}}, \boldsymbol{P}_3^{\boldsymbol{h}}, \boldsymbol{O}^{\boldsymbol{h}} \right\}$$

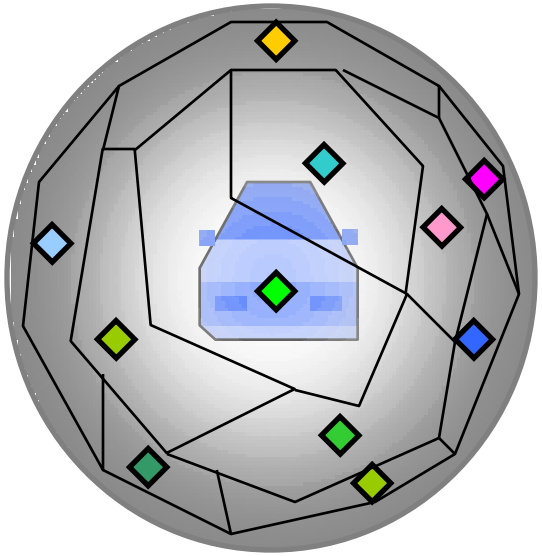
$$\tau : \boldsymbol{I}^{\boldsymbol{k}} \rightarrow \left\{ \boldsymbol{P}_1^{\boldsymbol{k}}, \boldsymbol{P}_2^{\boldsymbol{k}}, \boldsymbol{P}_3^{\boldsymbol{k}}, \boldsymbol{O}^{\boldsymbol{k}} \right\}$$

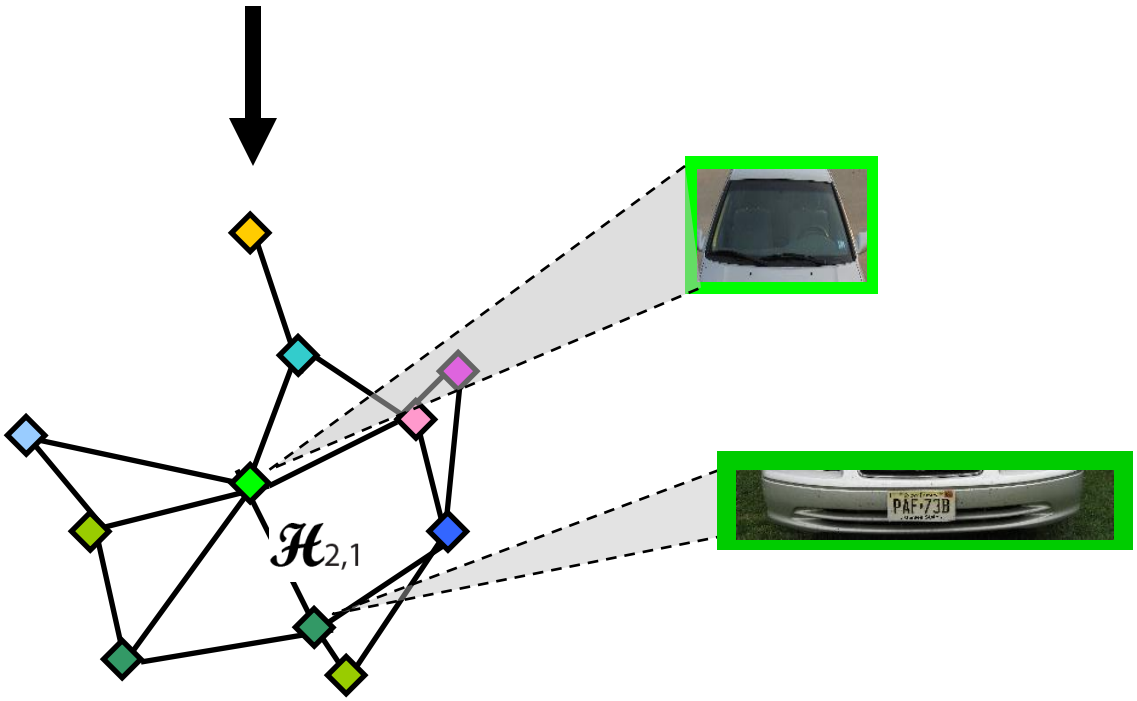
Canonical parts

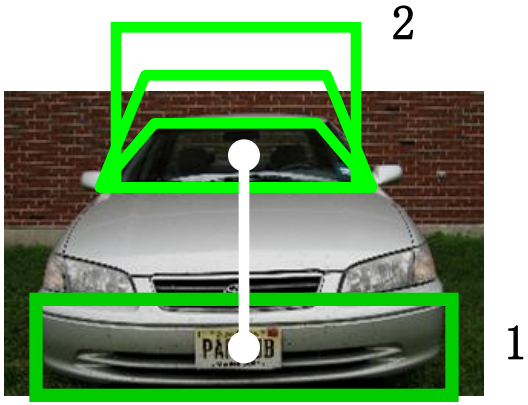




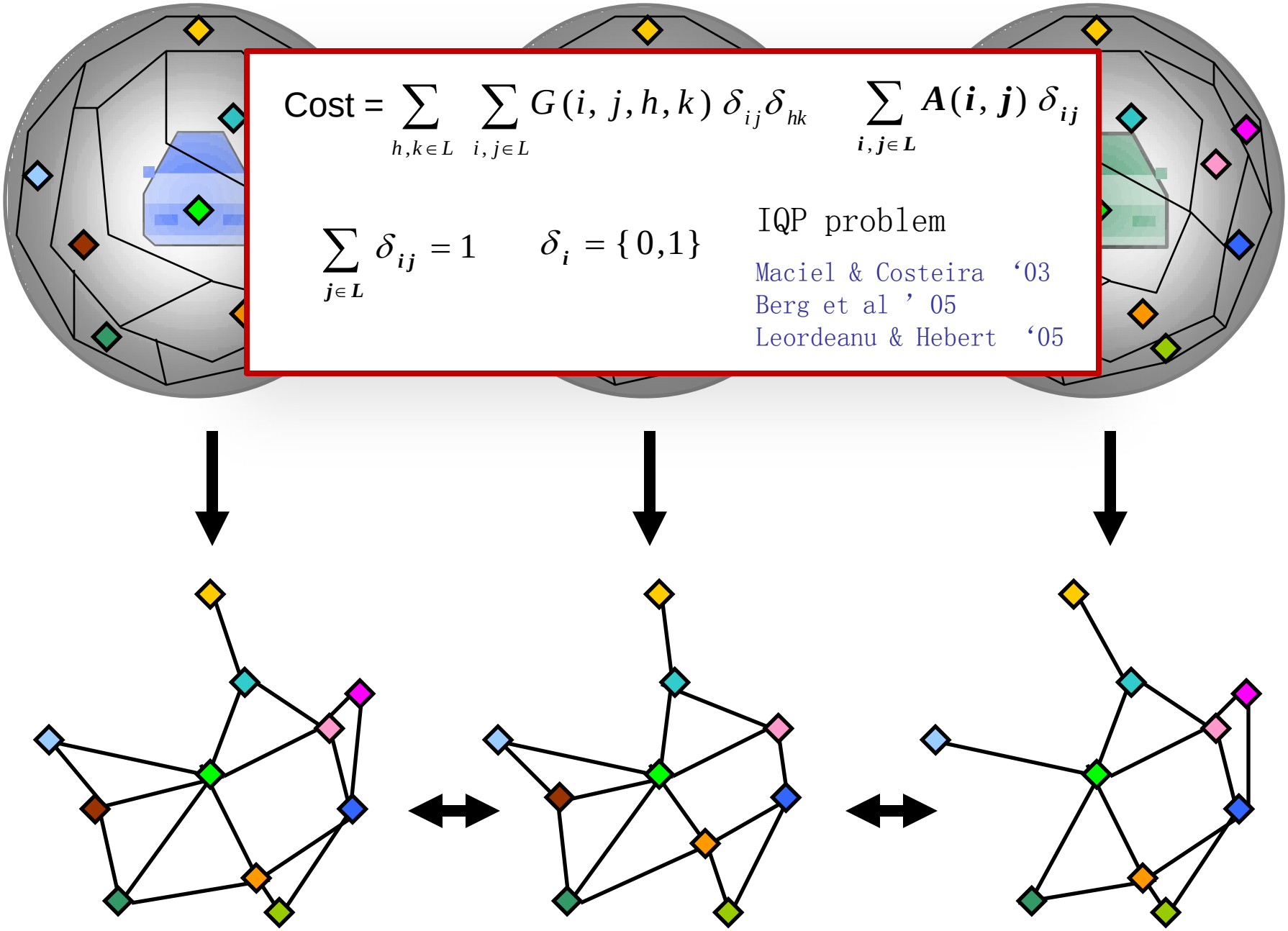
Linkage structure



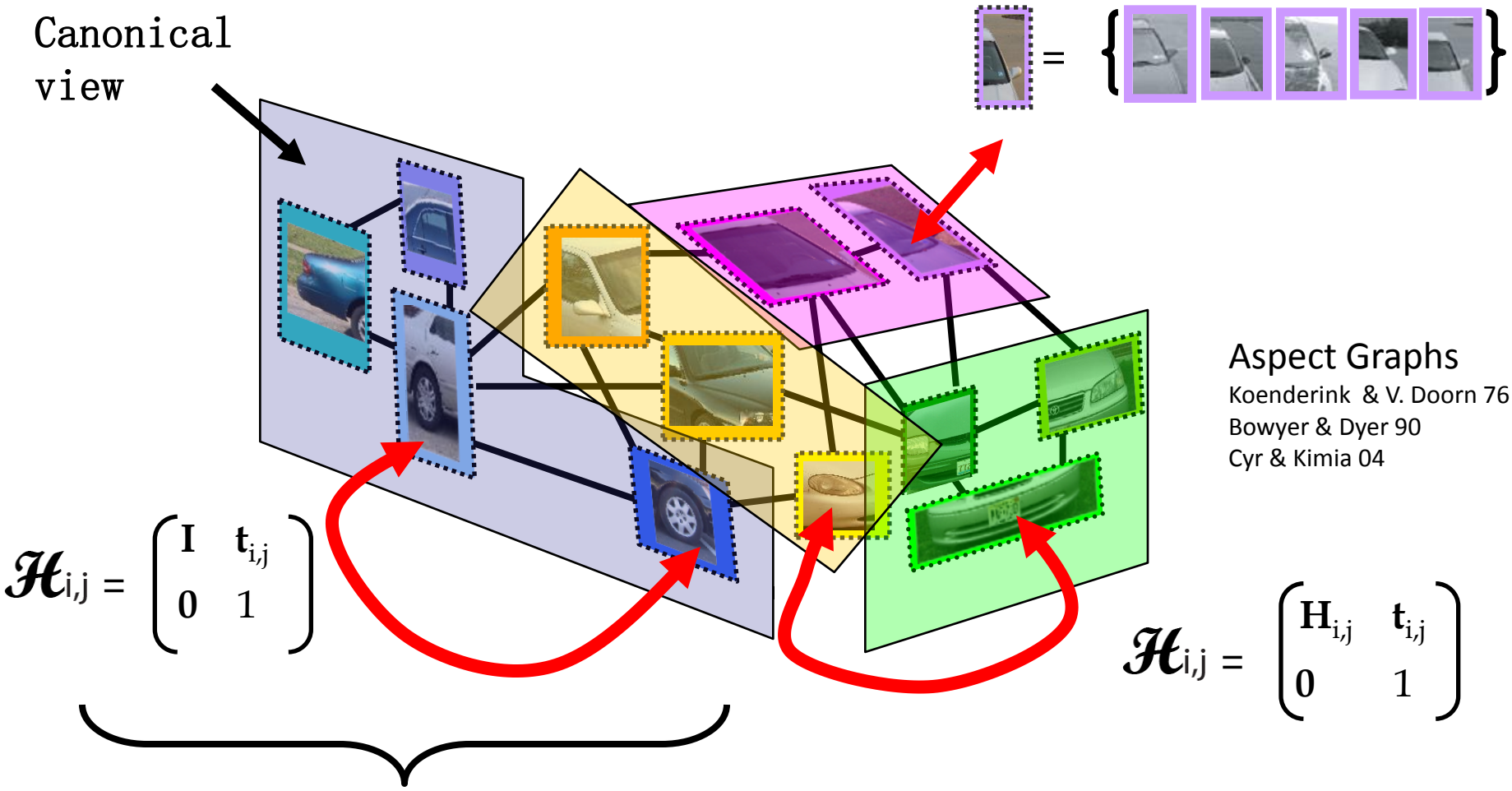




$$\mathcal{H}_{2,1} = \begin{bmatrix} \mathbf{H}_{2,1} & \mathbf{t}_{2,1} \\ \mathbf{0} & 1 \end{bmatrix}$$


Cost = $\sum_{h,k \in L} \sum_{i,j \in L} G(i,j,h,k)\, \delta_{ij} \delta_{hk}$ $\sum_{i,j \in L} A(i,j)\, \delta_{ij}$
$\sum_{j \in L} \delta_{ij} = 1$ $\delta_i = \{0,1\}$
IQP problem
Maciel & Costeira '03
Berg et al ' 05
Leordeanu & Hebert '05

Category Model

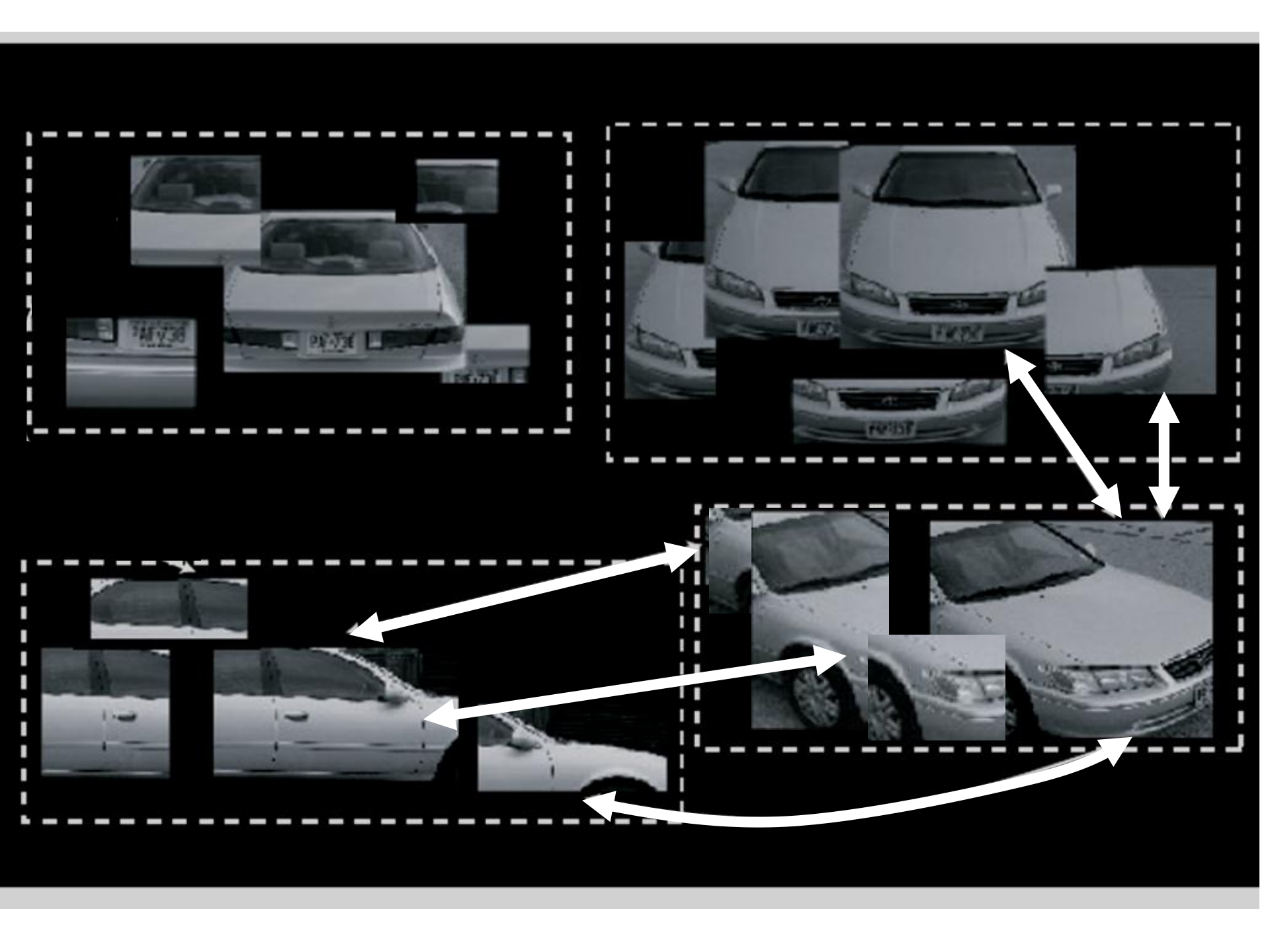


Aspect Graphs

Koenderink & V. Doorn 76

Bowyer & Dyer 90

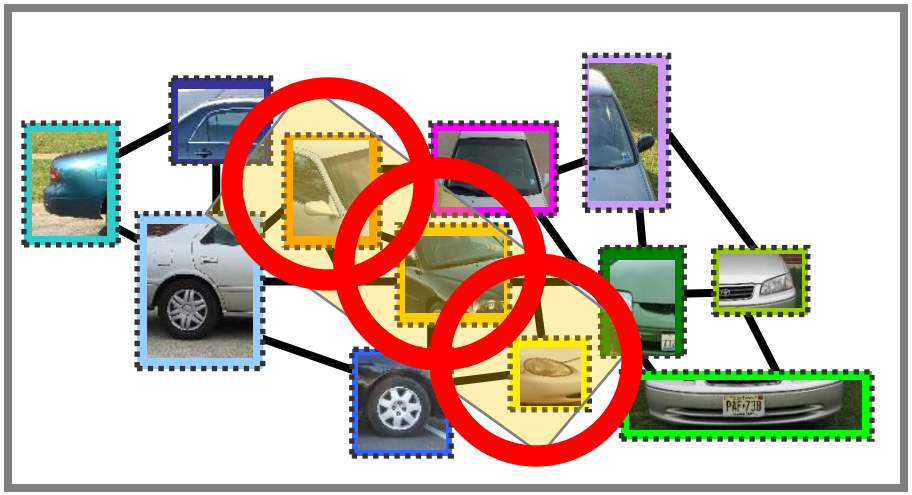
Cyr & Kimia 04



Object Recognition

Query image

model

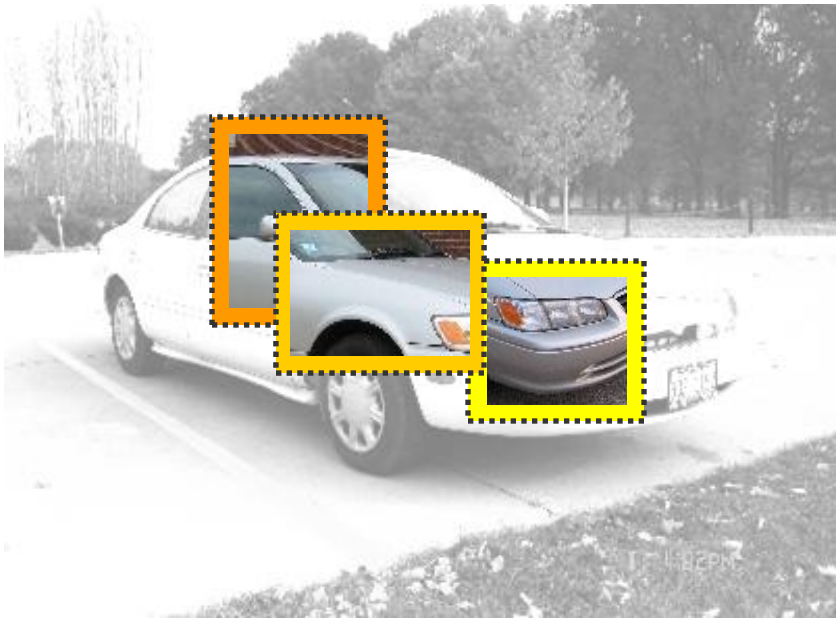


Algorithm

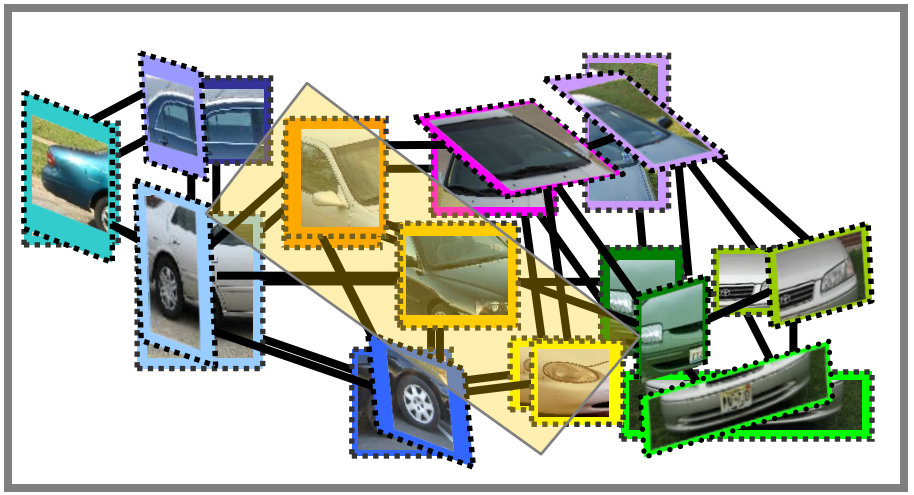
1. Find hypotheses of canonical parts consistent with a given pose

Object Recognition

Query image



model

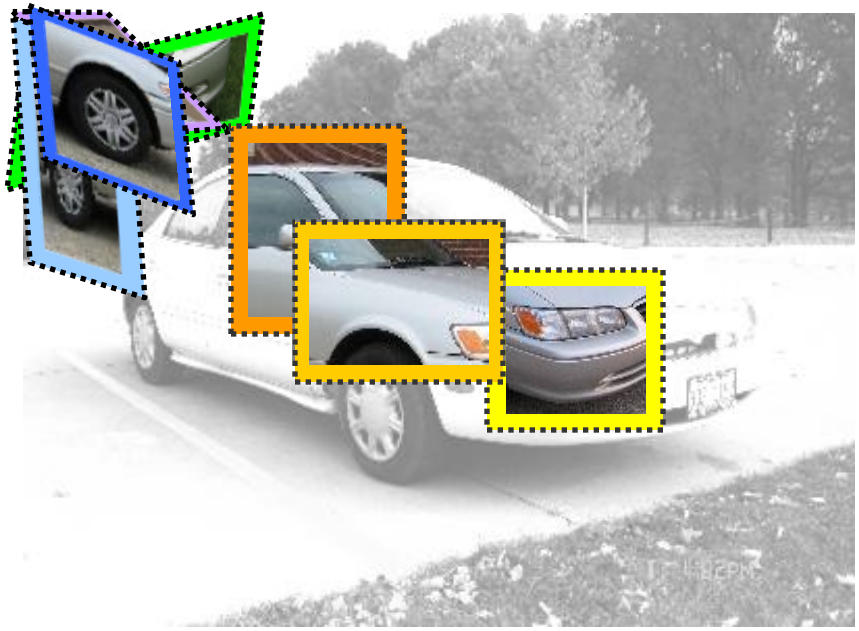


Algorithm

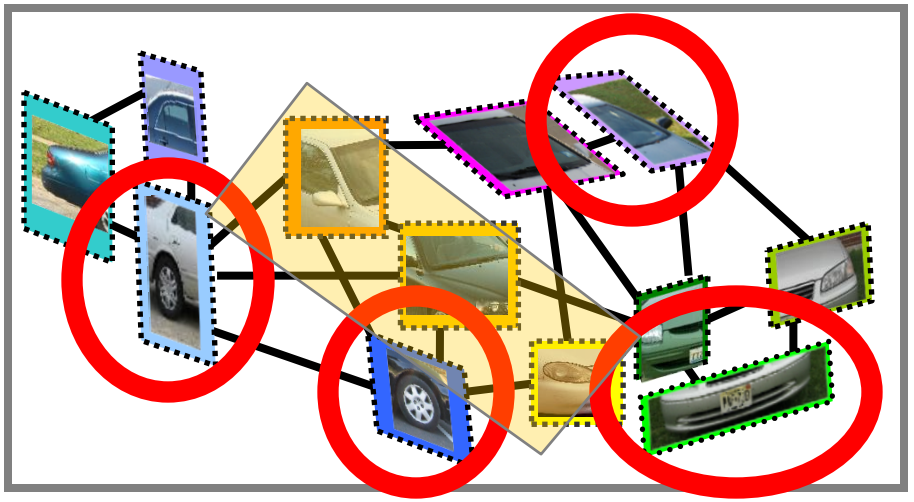
1. Find hypotheses of canonical parts consistent with a given pose
2. Infer position and pose of other canonical parts

Object Recognition

Query image



model

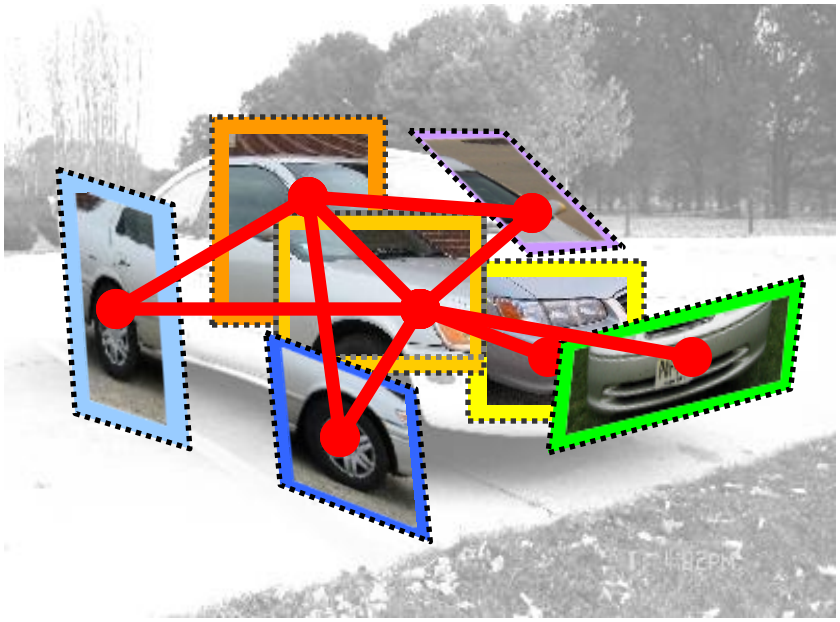


Algorithm

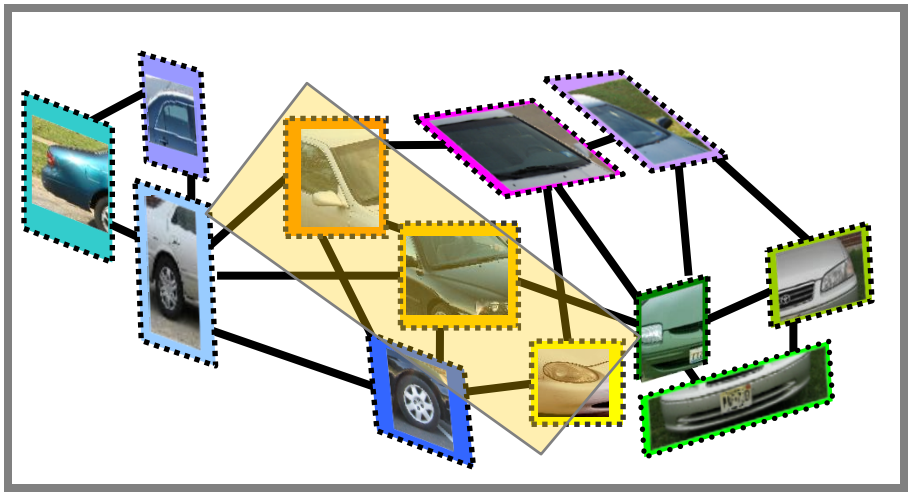
1. Find hypotheses of canonical parts consistent with a given pose
2. Infer position and pose of other canonical parts

Object Recognition

Query image



model

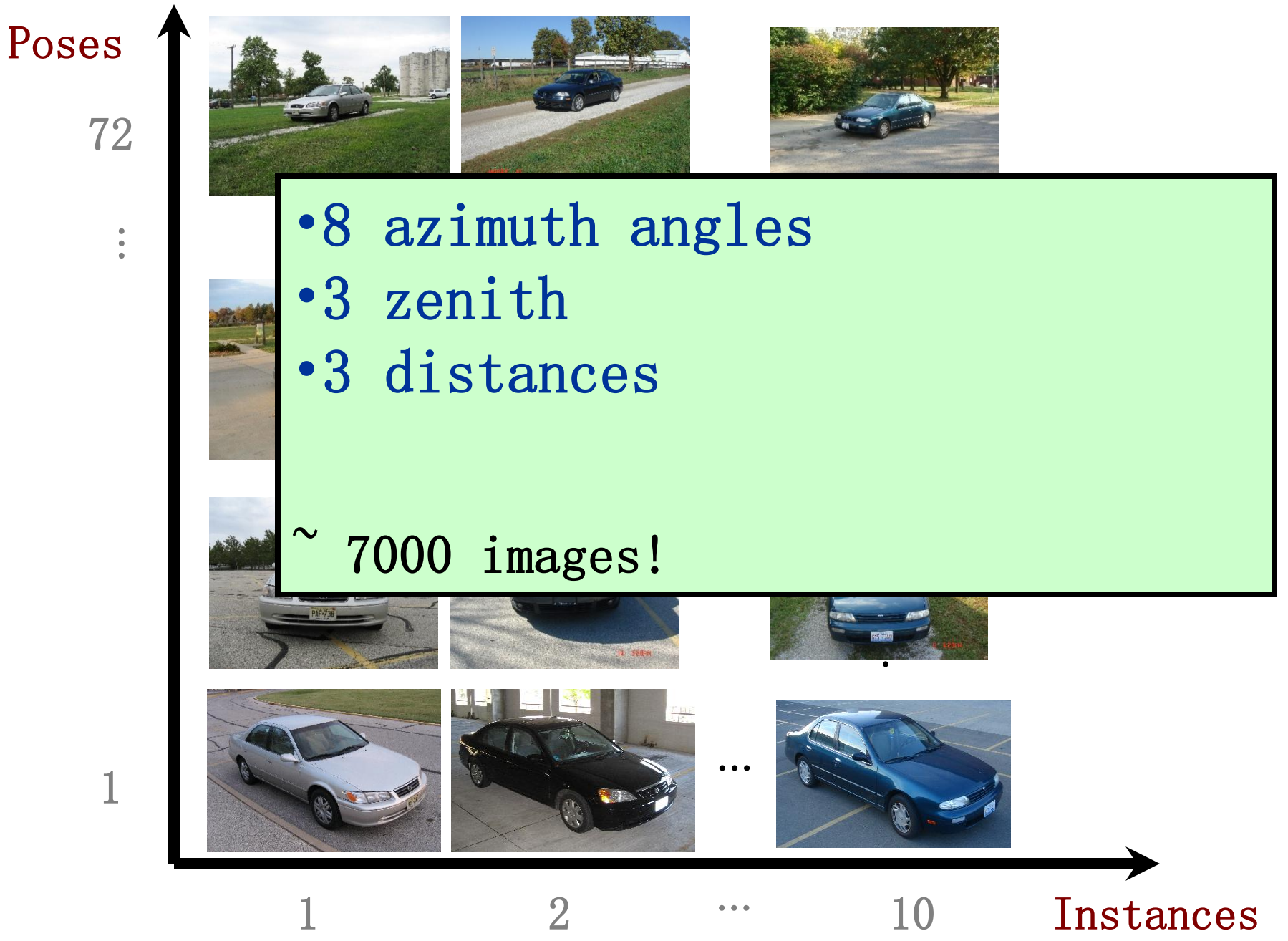


Algorithm

1. Find hypotheses of canonical parts consistent with a given pose
2. Infer position and pose of other canonical parts
3. Optimize over **E**, **G** and **s** to find best combination of hypothesis

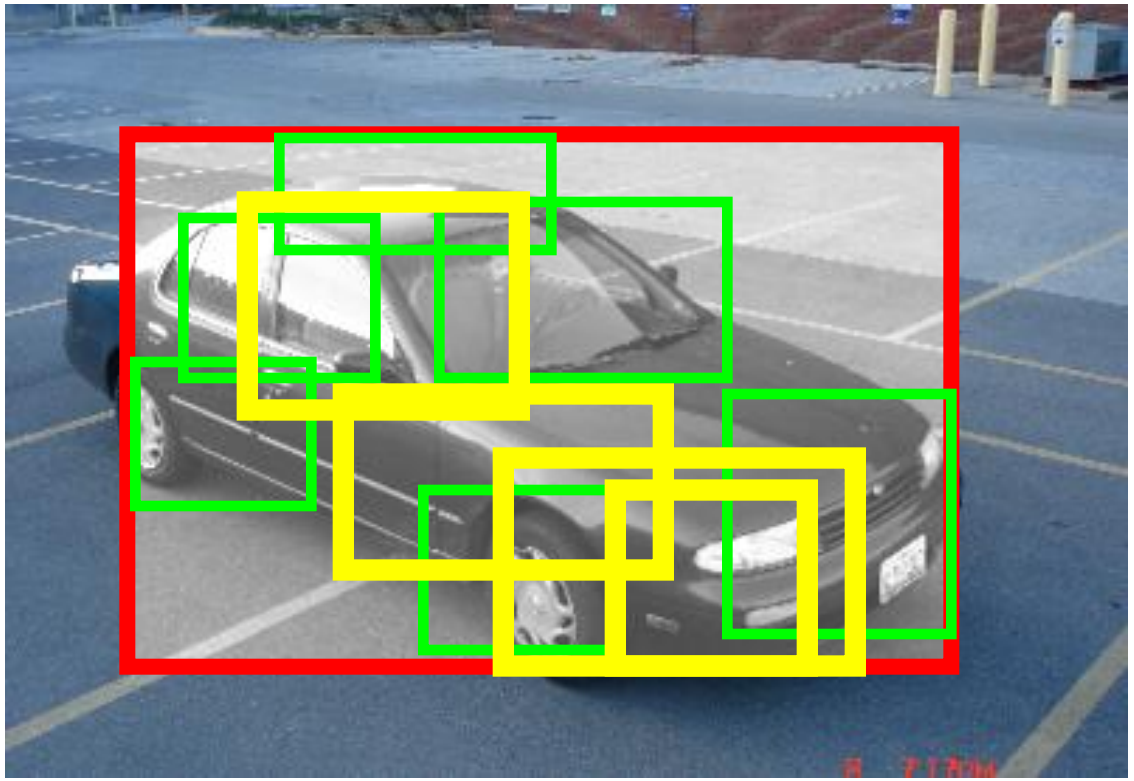
$\rightarrow$ error

3D object class dataset


Poses
72
1
•8 azimuth angles
•3 zenith
•3 distances
~ 7000 images!
1
2
10
Instances

Examples

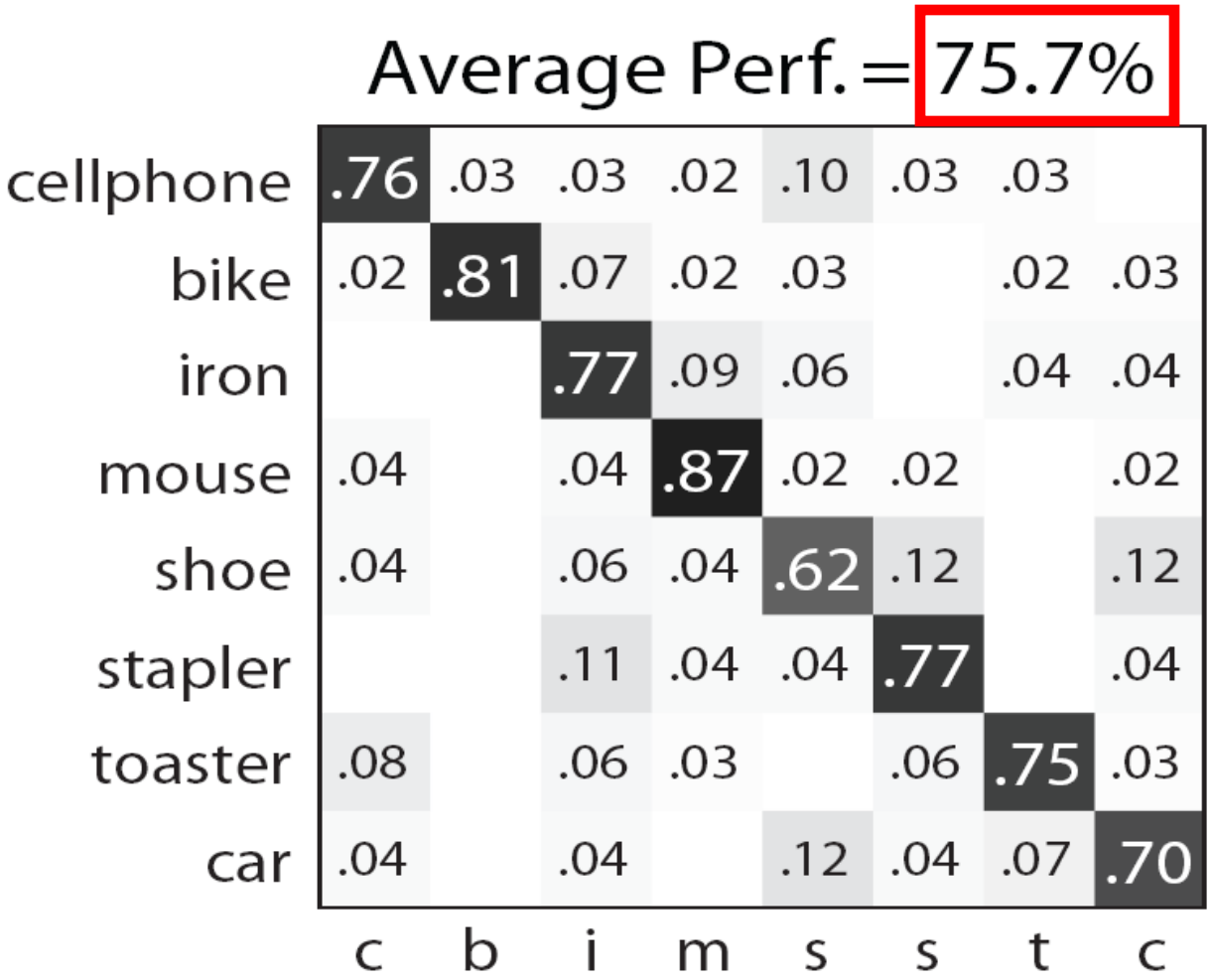
Category: car
Azimuth = 45°
Zenith = 30°
Distance = close



Examples

Category: mouse
Azimuth = 315°
Zenith = 0°
Distance = medium

Classification accuracy



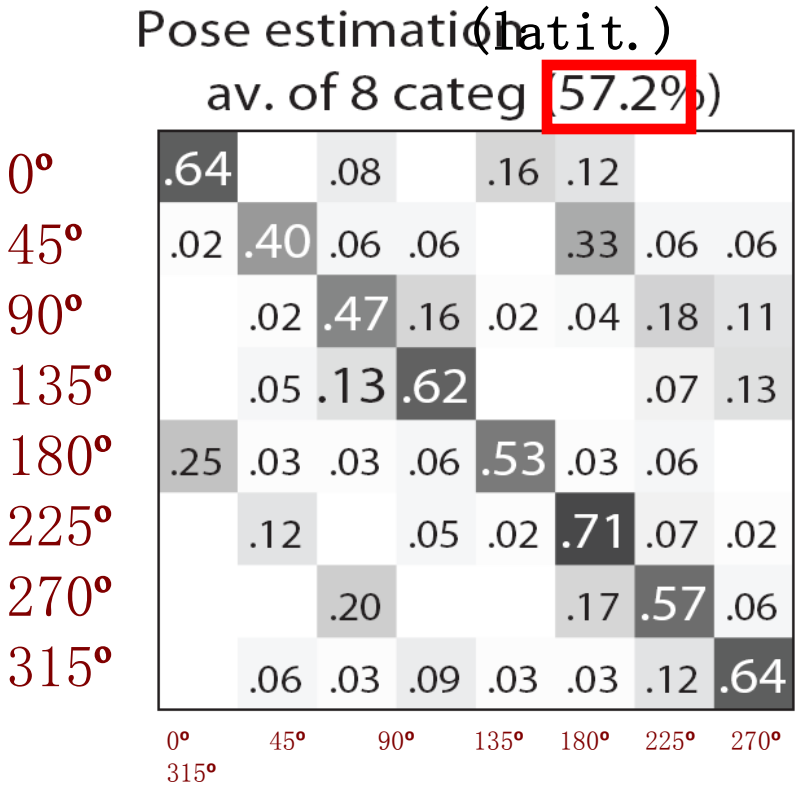
Average accuracy in classifying 1 out 8 categories

Random=12%

Failure example

Category: car
Azimuth = 225°
Zenith = 30°
Distance = close

Pose estimation accuracy



Summary

	Single view	Mixture / Multi-view	Sav. et al, 07
View point invariant	X	√ Category	√
No supervision	√	X View point	X → √ category ; all views all instances available
# Categories	~300	2	8
	√		

- lack of a coherent methodology for learning parameters
- need of multi-view observations of the same object instance
- no generative model for robust learning and recognition

Part-based Multi-view Models

Savarese, Fei-Fei, ICCV 07

Savarese, Fei-Fei, ECCV 08

Sun, Su, Savarese, Fei-Fei, CVPR 09

Su, Sun, Fei-Fei, Savarese, ICCV 09

Min Sun
University of Michigan, USA

Hao Su
Beihang University, China

Fei-Fei Li.
Stanford U, USA

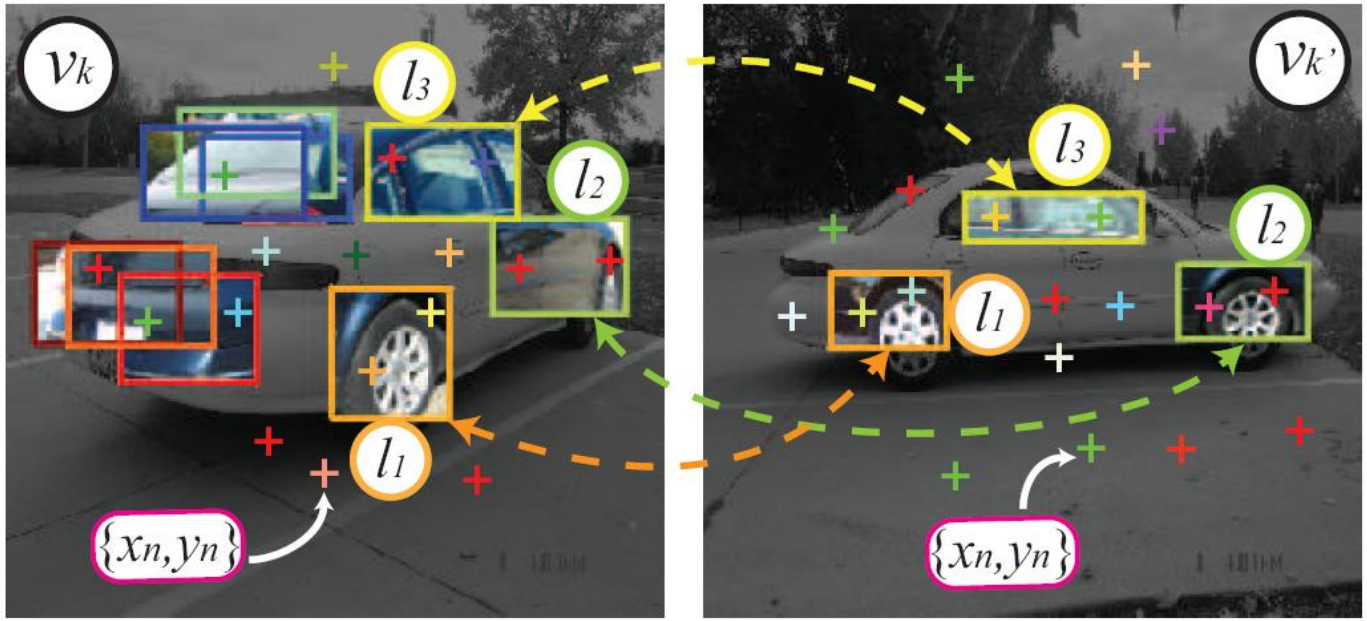
Part-based Multi-view Models

Savarese, Fei-Fei, ICCV 07

Savarese, Fei-Fei, ECCV 08

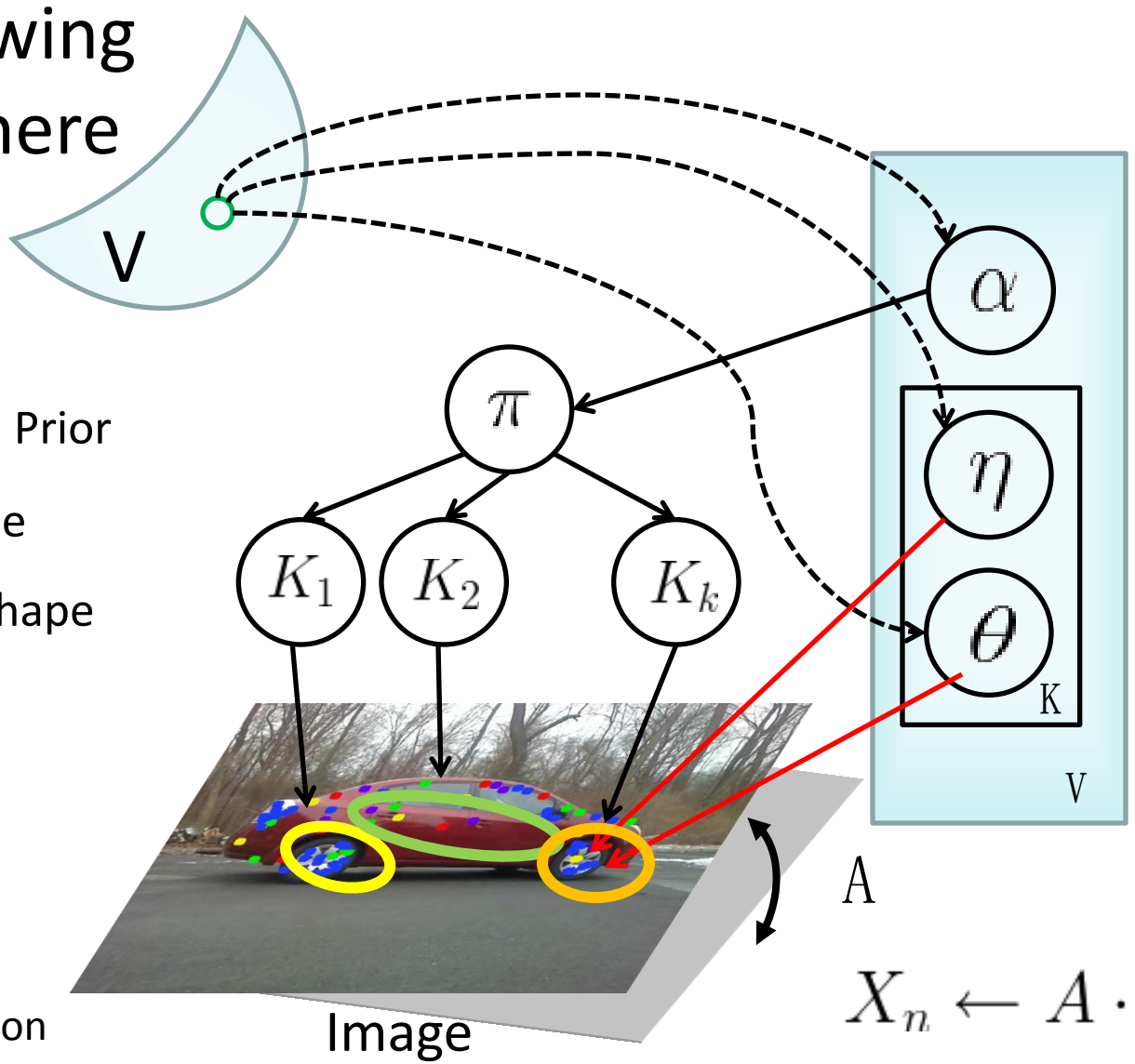
Sun, Su, Savarese, Fei-Fei, CVPR 09

Su, Sun, Fei-Fei, Savarese, ICCV 09



- Probabilistic generative part-based model
- Dense multi-view representation on the viewing sphere

Part-based generative model



α = Part Proportion Prior

η = Part Appearance

θ = Part Location/shape

$$\pi \sim Dir(\alpha)$$

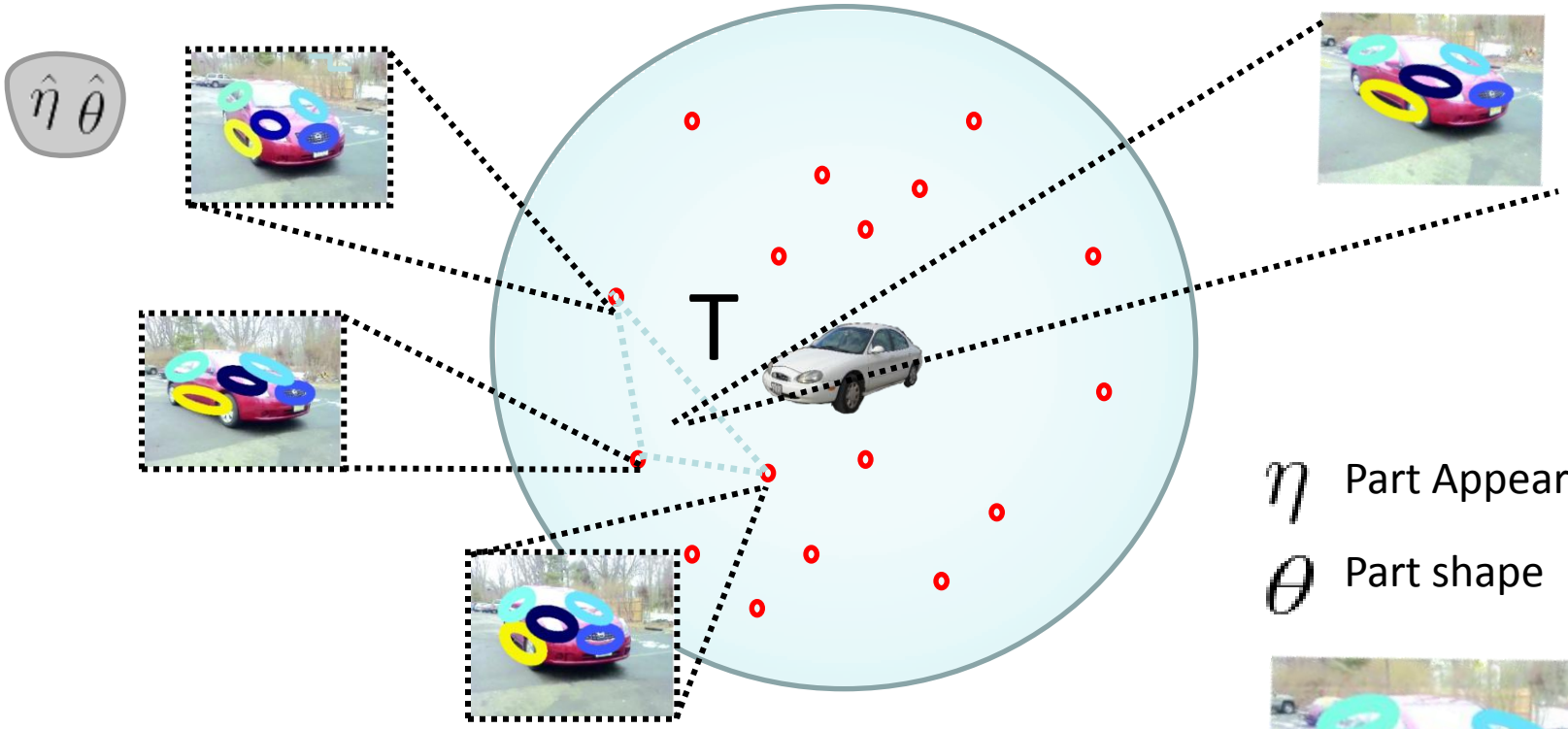
$$Y_n \sim Mult(\eta)$$

$$X_n \sim N(\theta)$$

Y=Codeword, X=Location

$$X_n \leftarrow A \cdot X$$

Dense representation on view-sphere



η Part Appearance

θ Part shape

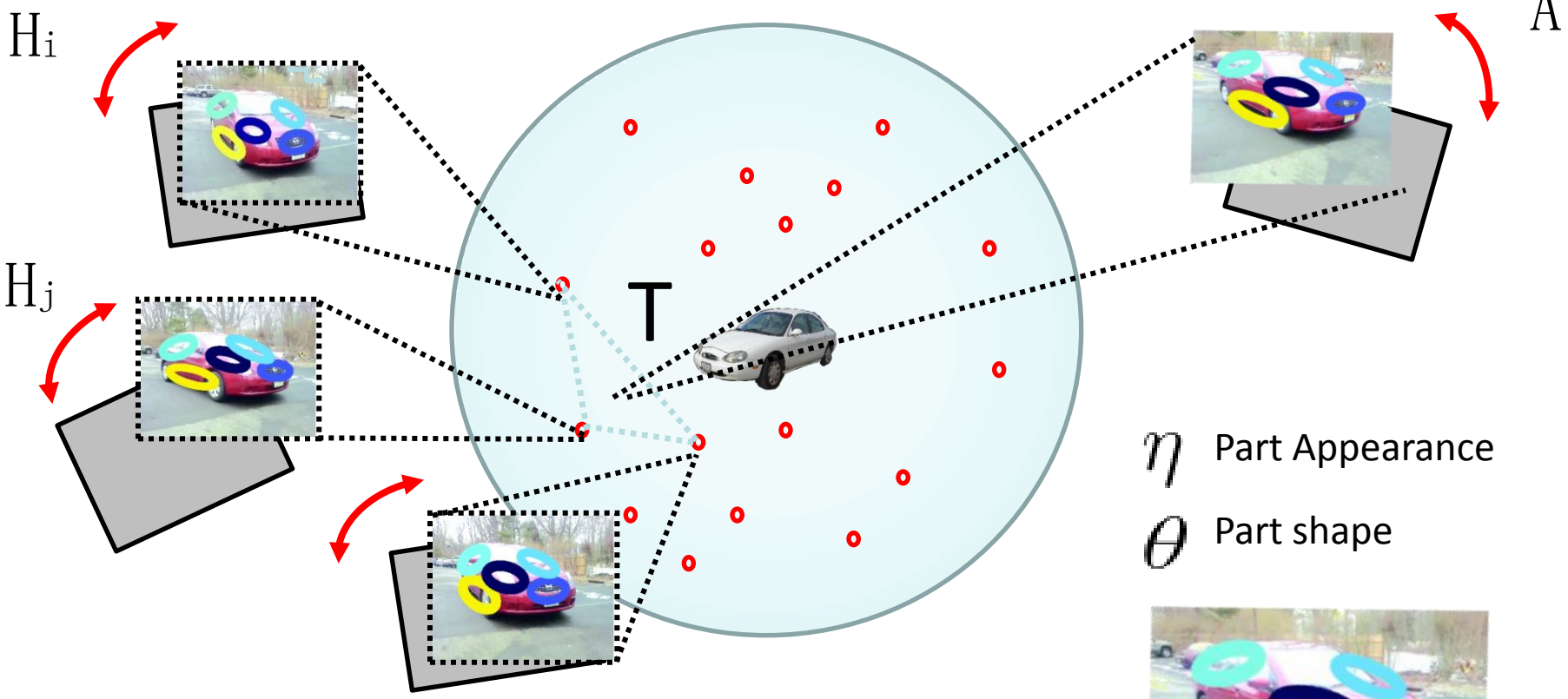
Constraints based on view morphing geometry

$$m = \sum_{g=1}^{3} \hat{m}_{Tk}^{g} \cdot S_g$$

Seitz &Dyer SIGGRAPH 96
Xiao & Shah CVIU '04

$$\theta = (m, \Sigma)$$

Dense representation on view-sphere



η Part Appearance

θ Part shape

- Pre-warping transformations H
- Post-warping transformation A

$$\theta = (m, \Sigma)$$

Dense representation on view-sphere

Joint probability of the model:

$$P(X, Y, T, S, K, \pi) = P(T|\phi)P(\pi|\alpha_T)P(S|\beta)$$
$$\prod_n^N \{P(x_n|\theta_{TK_n}(S), A)P(y_n|\eta_{TK_n}(S))P(K_n|\pi)\}$$

Observable variables: X, Y, T, S

Latent variables: K, π + relevant priors

Learning

Variational EM

E step:

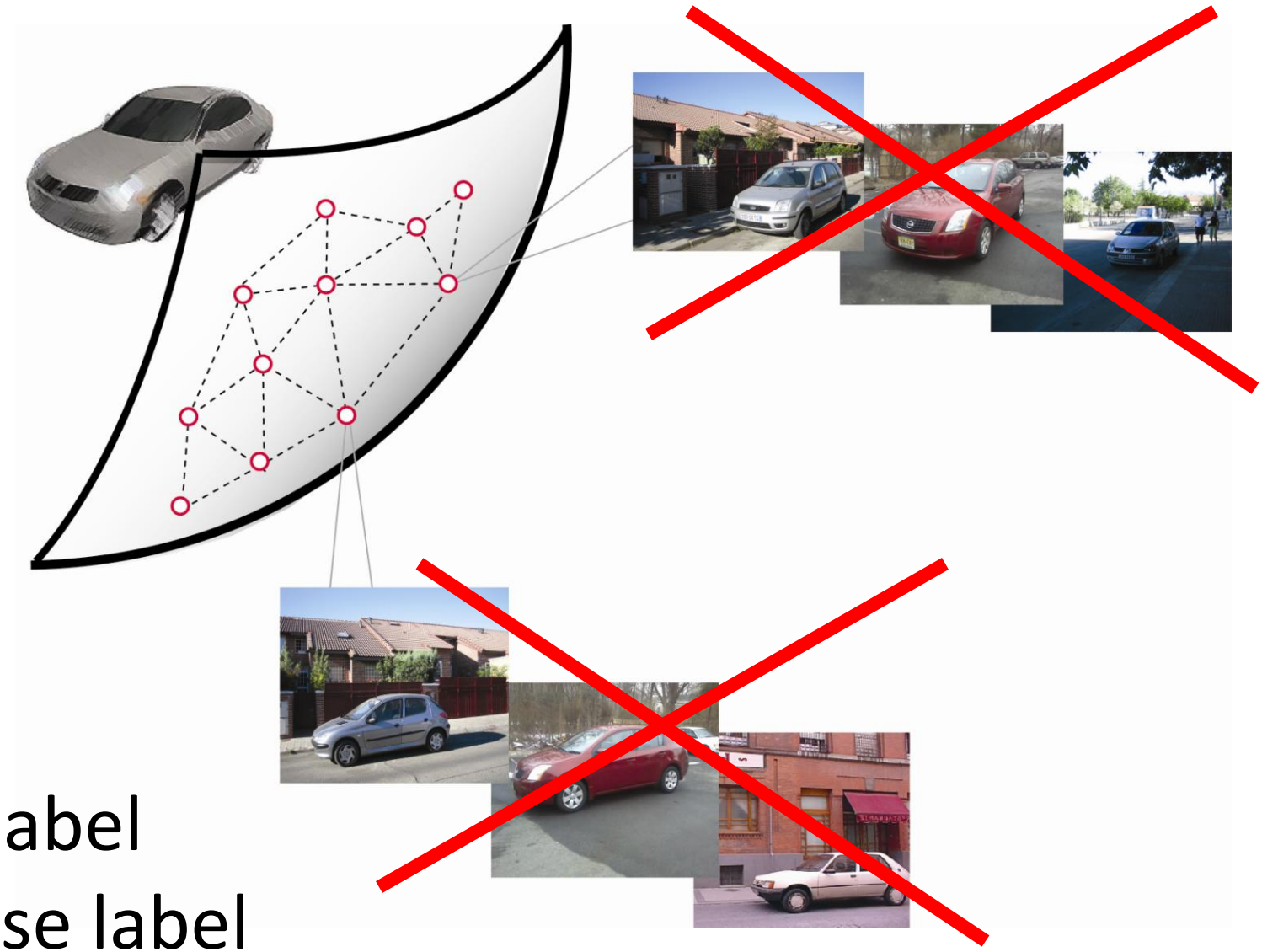
- update hidden part assignments π
  part proportion K

M step:

- update part proportion prior α
  appearance η
  Location/shape θ

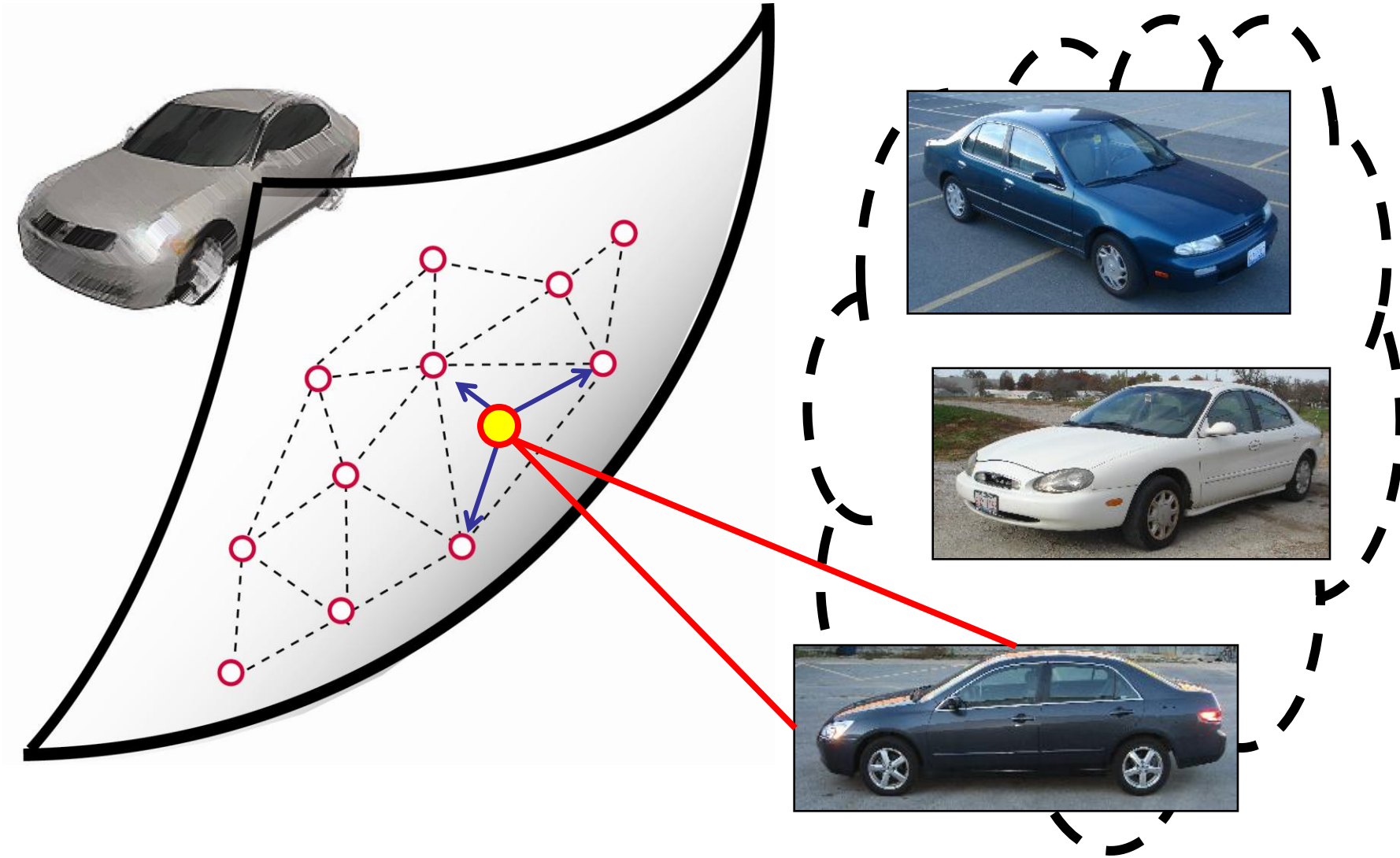
- weakly supervised
- incremental (training image)

Weakly supervised

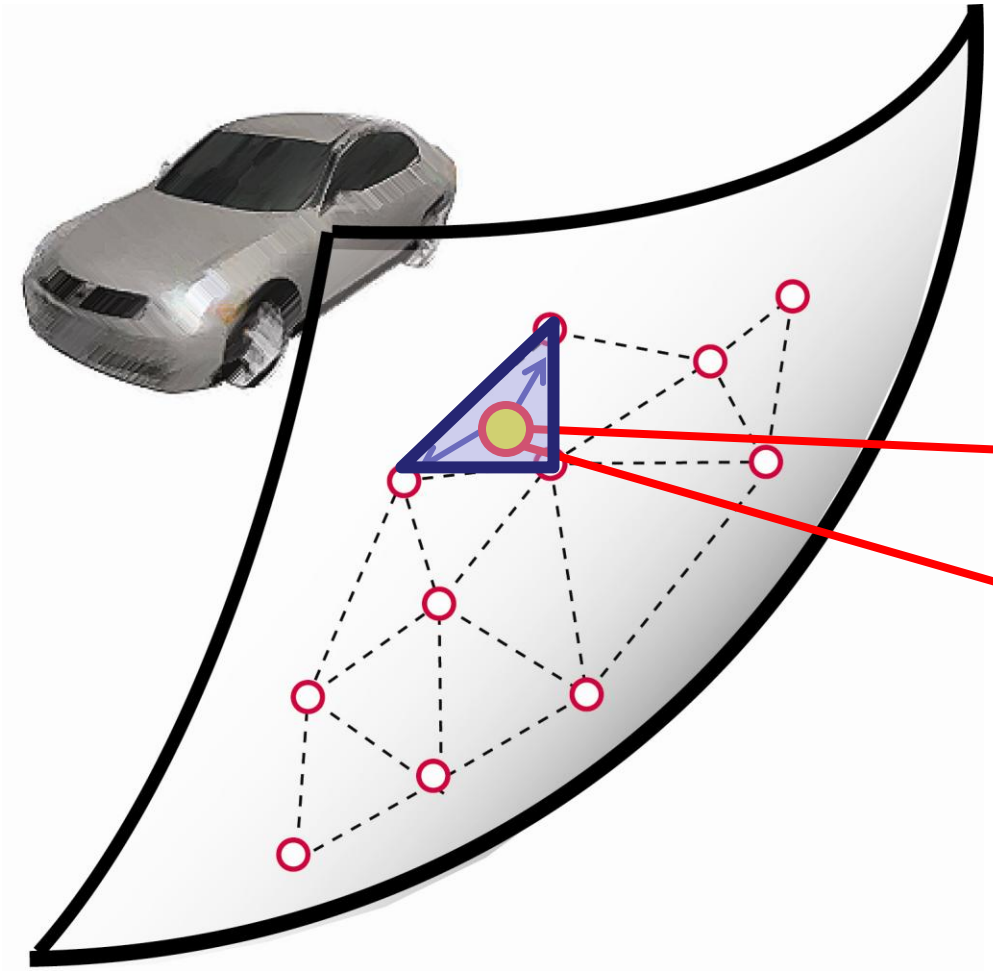


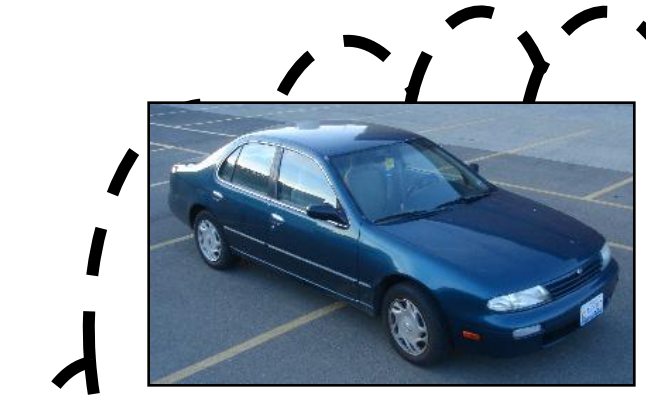
- Class label
- No pose label

Incremental learning



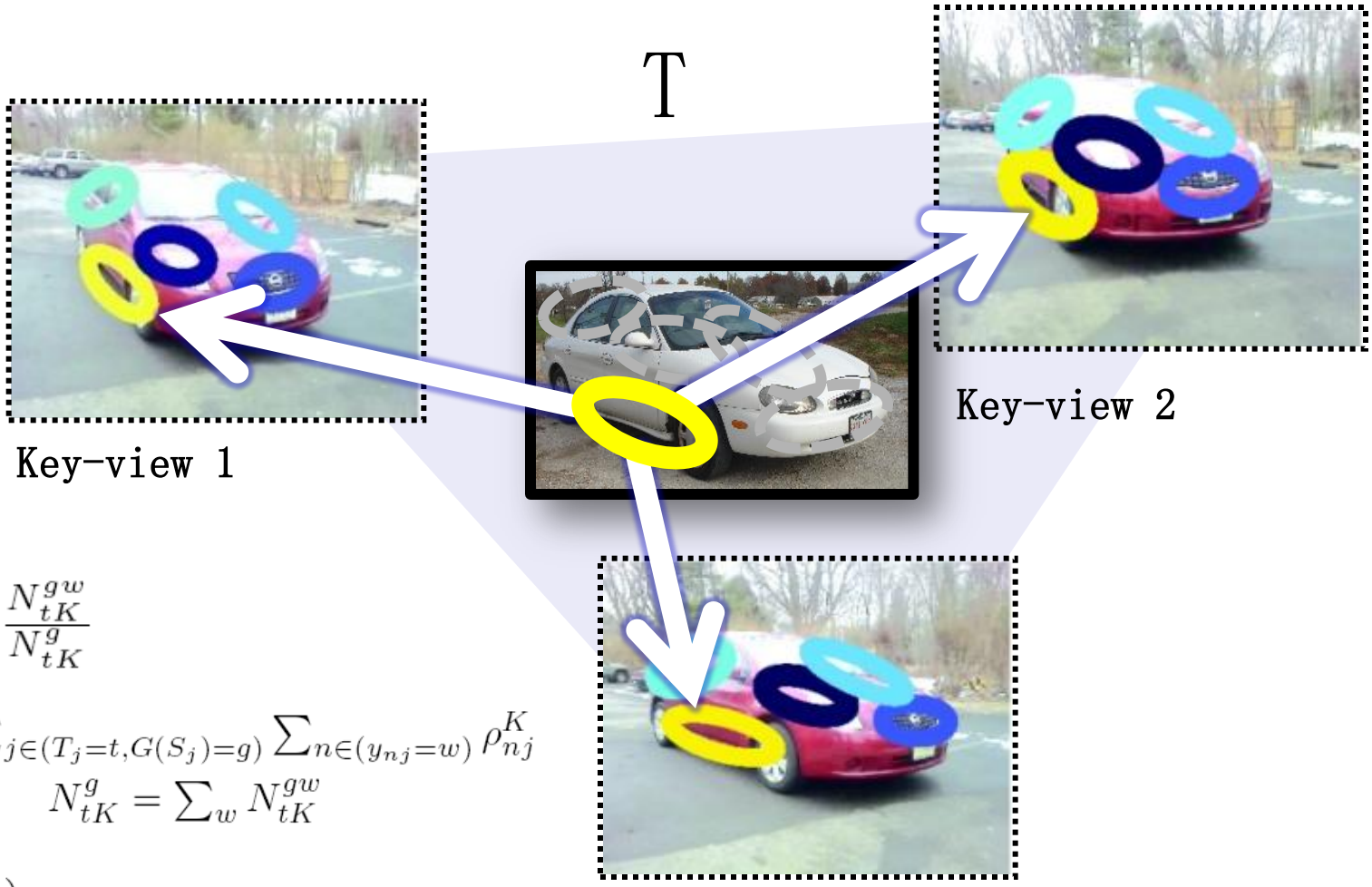
Incremental learning





No need for observations of the same object instance from multiple view

Incremental learning



$$\hat{\eta}_{tK}^{gw} = \frac{N_{tK}^{gw}}{N_{tK}^{g}}$$

$$N_{tK}^{gw} = \sum_{j \in (T_j = t, G(S_j) = g)} \sum_{n \in (y_{nj} = w)} \rho_{nj}^{K}$$

$$N_{tK}^{g} = \sum_{w} N_{tK}^{gw}$$

$$N_{T_j K}^{G(S_j) w} + \sum_{n \in (y_{nj} = w)} \rho_{nj}^{K}$$

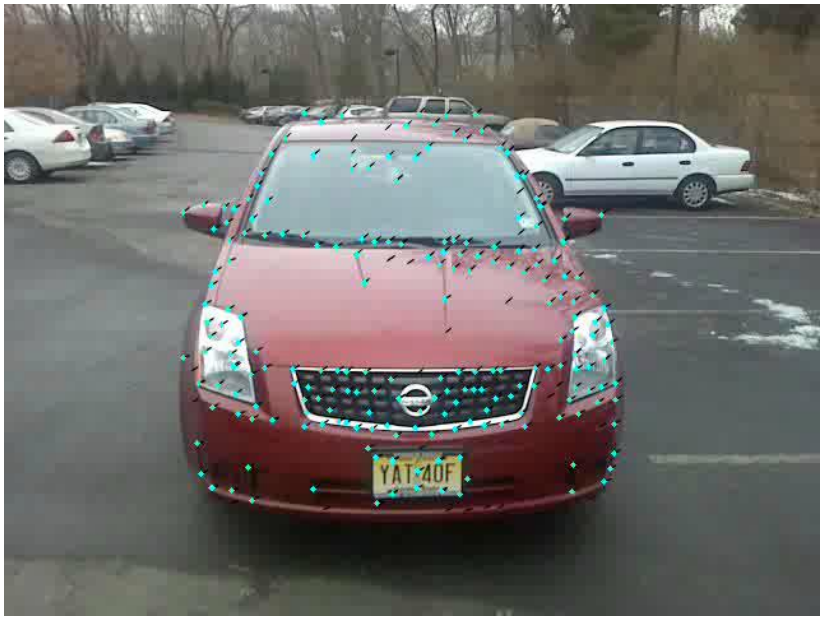
key ingredients for weakly supervised learning

Initialization

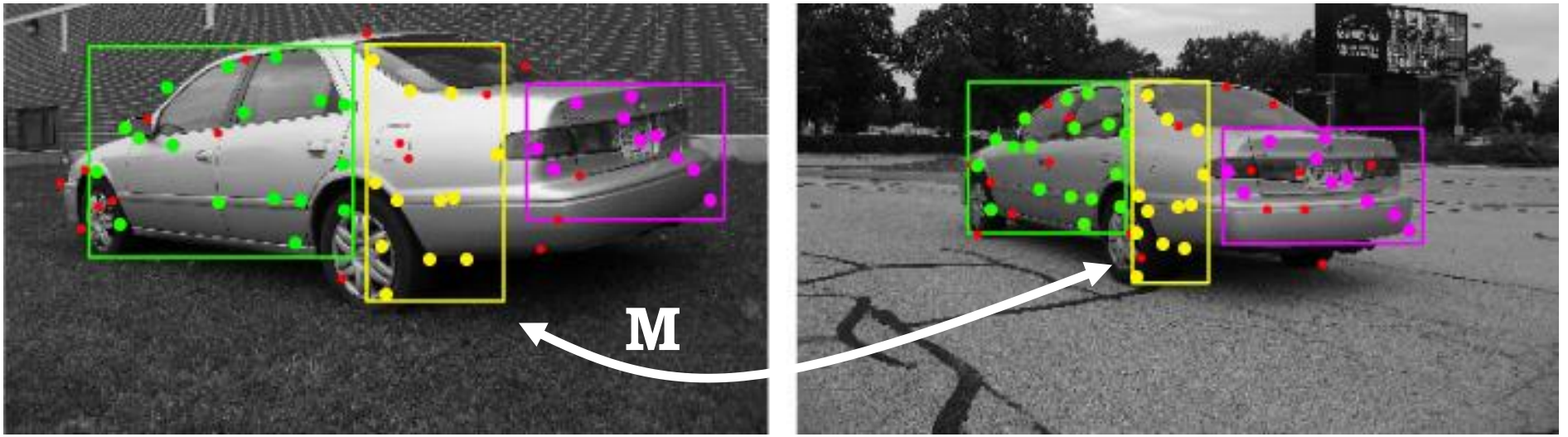
Constraints

- Across Triangle
- Within Triangle

Initialization



Initialization



$$\pi : \boldsymbol{I}^{h} \rightarrow \left\{ \boldsymbol{P}_1^{h}, \boldsymbol{P}_2^{h}, \boldsymbol{P}_3^{h}, \boldsymbol{O}^{h} \right\}$$

$$\tau : \boldsymbol{I}^{k} \rightarrow \left\{ \boldsymbol{P}_1^{k}, \boldsymbol{P}_2^{k}, \boldsymbol{P}_3^{k}, \boldsymbol{O}^{k} \right\}$$

Sequential ransac
J-linkage

key ingredients for weakly supervised learning

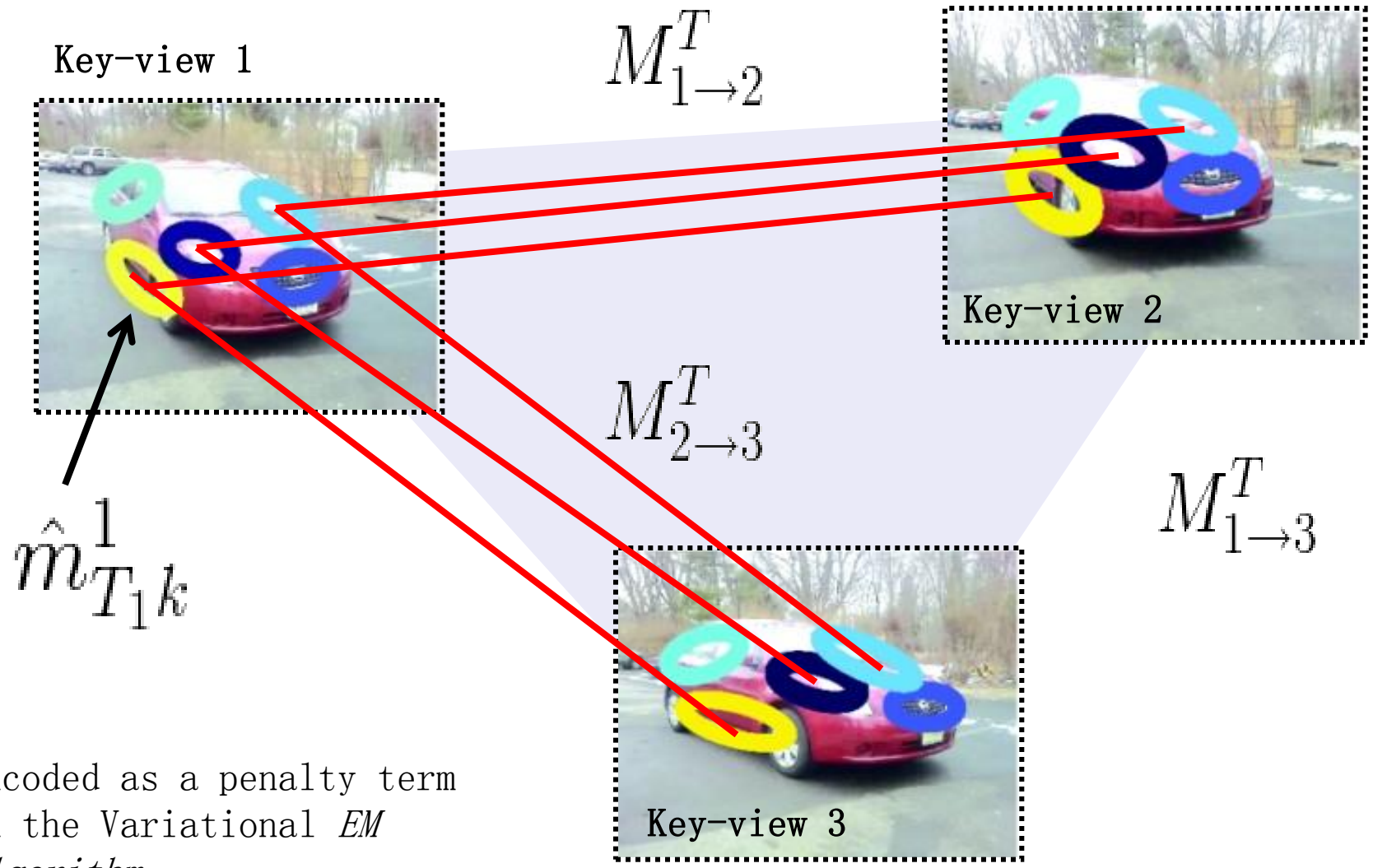
Initialization

Constraints

- Across Triangle
- Within Triangle

Constraints

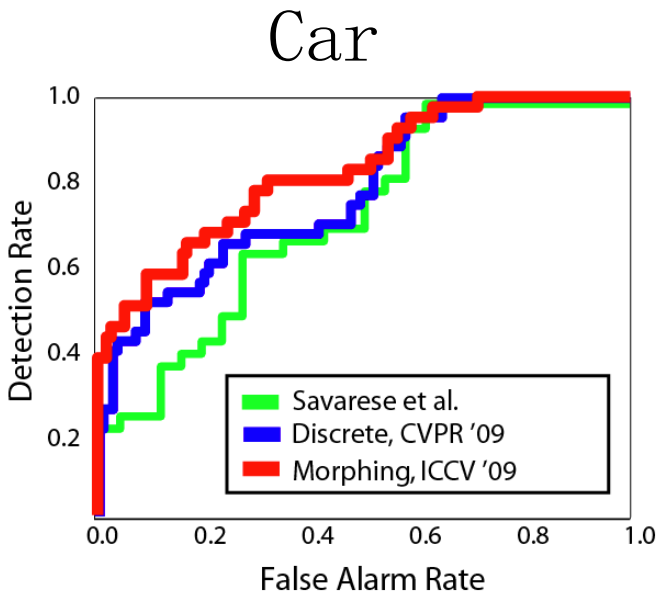
Within Triangle Constraints: $M^T_{i\to j} \cdot \hat{m}^i_{Tk} \approx \hat{m}^j_{Tk}$

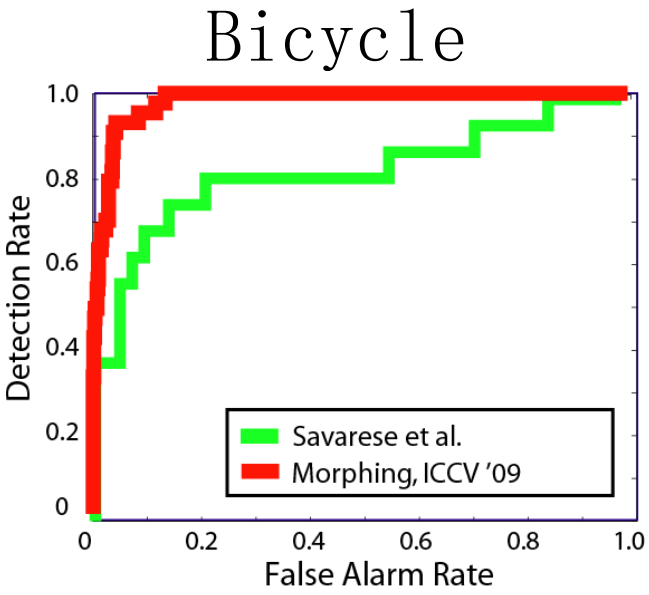


Encoded as a penalty term in the Variational *EM algorithm*

3D object class dataset

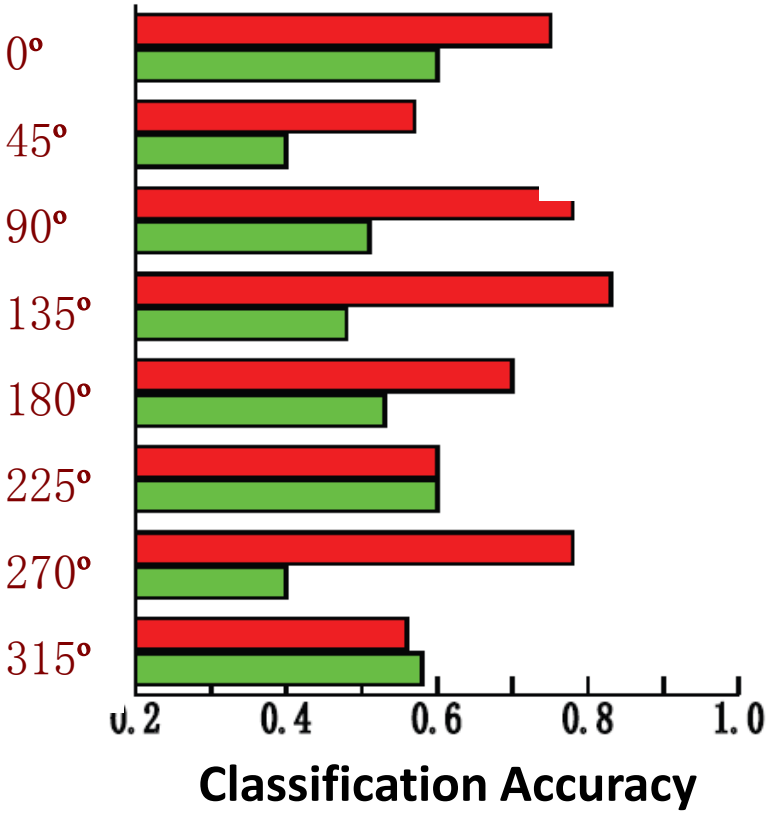
Detection





- Morphing model, 2009
- Savarese, & Fei-Fei ICCV '07

Viewpoint Classification: Car



Morphing model, 2009

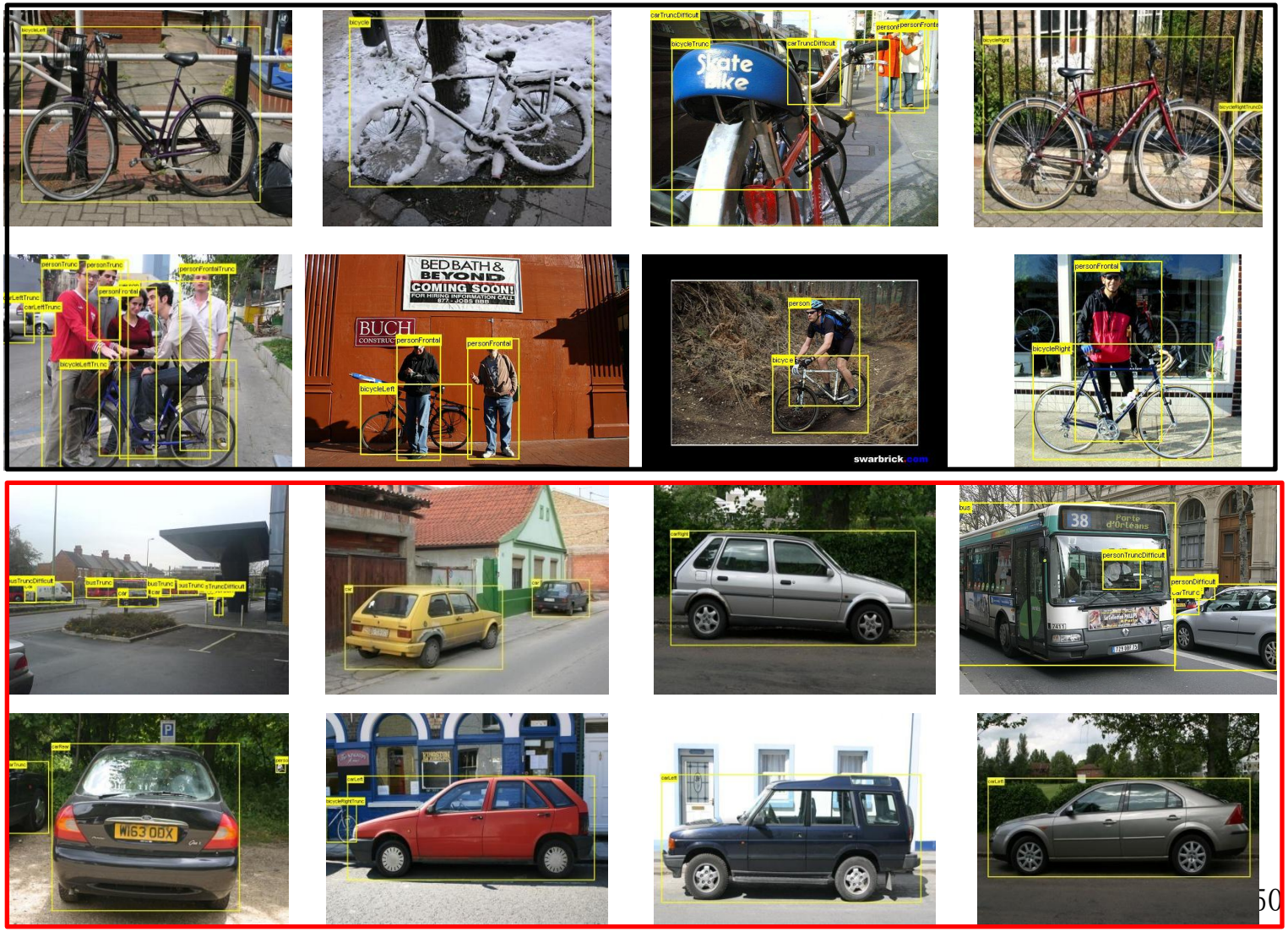
Savarese, & Fei-Fei ICCV '07

'06

Detection: Pascal 2006 dataset

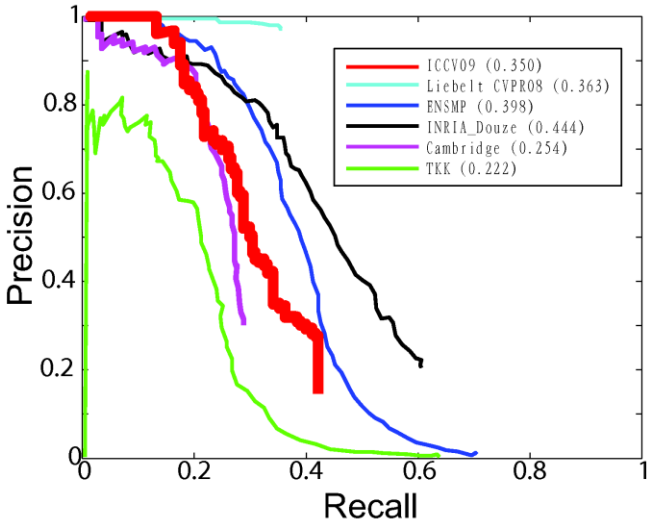
Bicycle

Car



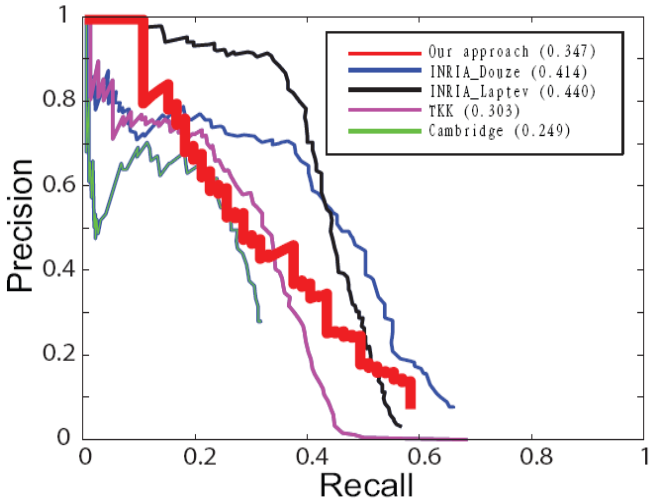
Detection: Pascal 2006 dataset

Car



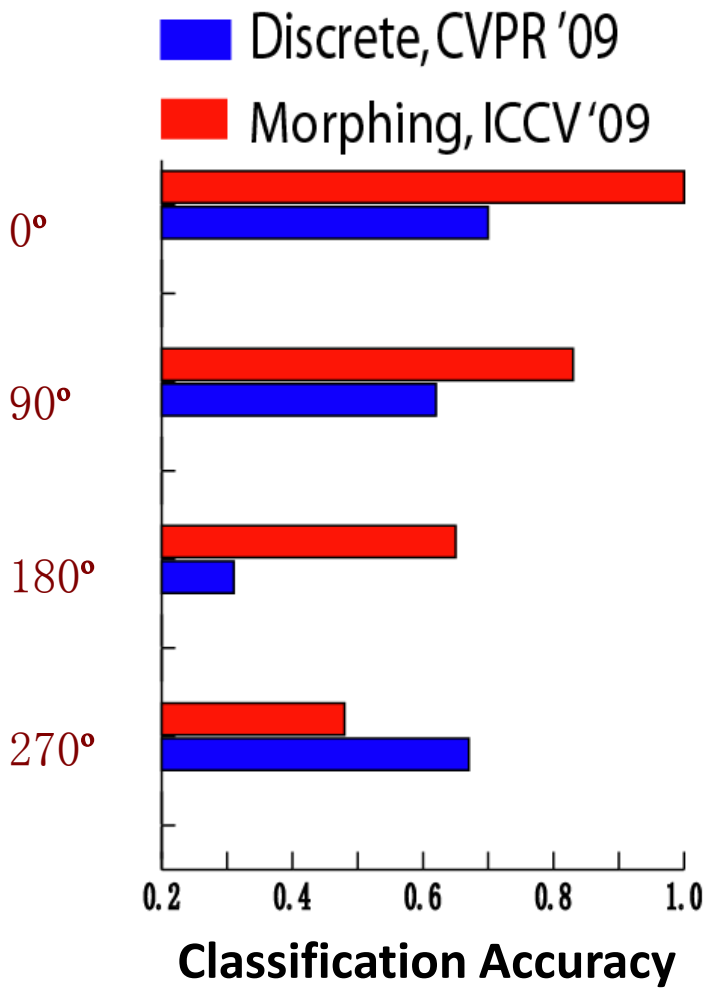
ICCV09 Morphing Model
0.35(average p)

Bicycle



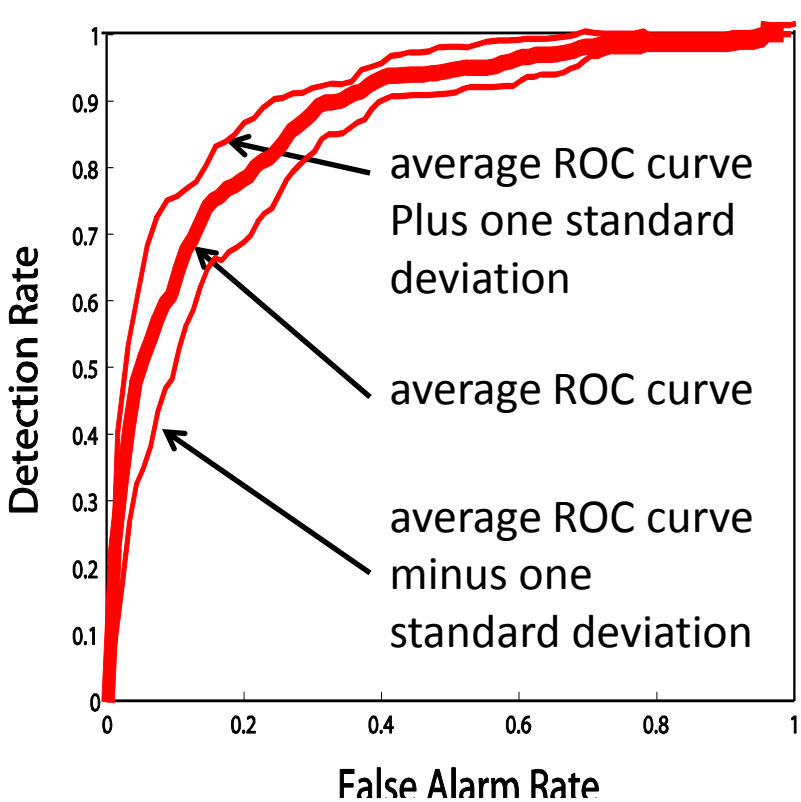
ICCV09 Morphing Model
0.347(average p)

Viewpoint Classification: Car- Pascal 2006 dataset



Pascal '06

Household Item Dataset: Detection



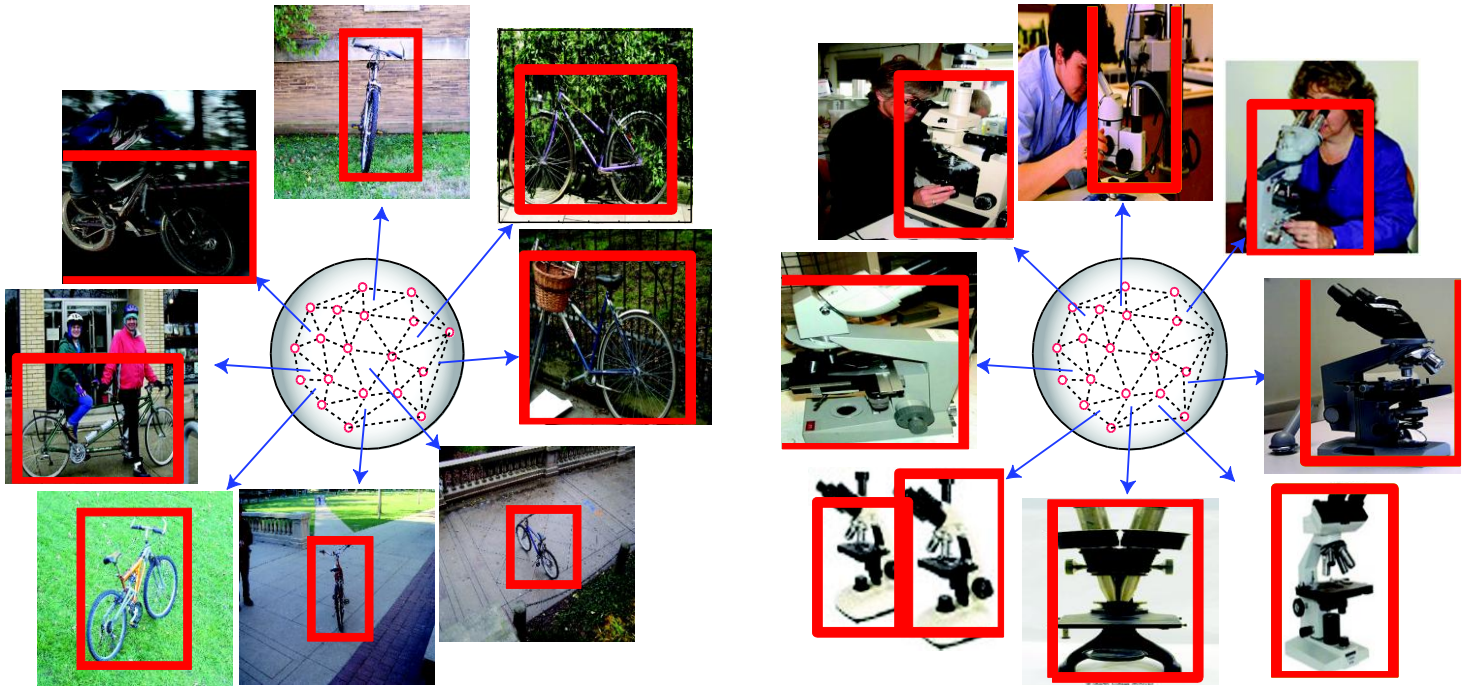
Object Class	AUC
Sewing Machine	98.1
Microscope	87.9
Travel Iron	88.1
Swivel Chair	91.2
Calculator	97.2
Flashlight	87.1
Teapot	86.4
Watch	84.9
All	90.1

Household Item Dataset: 8-Viewpoint Classification

Object Class	Accuracy
Sewing Machine	71.4
Microscope	63.9
Travel Iron	73.5
Swivel Chair	58.6
Calculator	69.2
Flashlight	68.4
Teapot	60.0
Watch	61.9
All	70.2

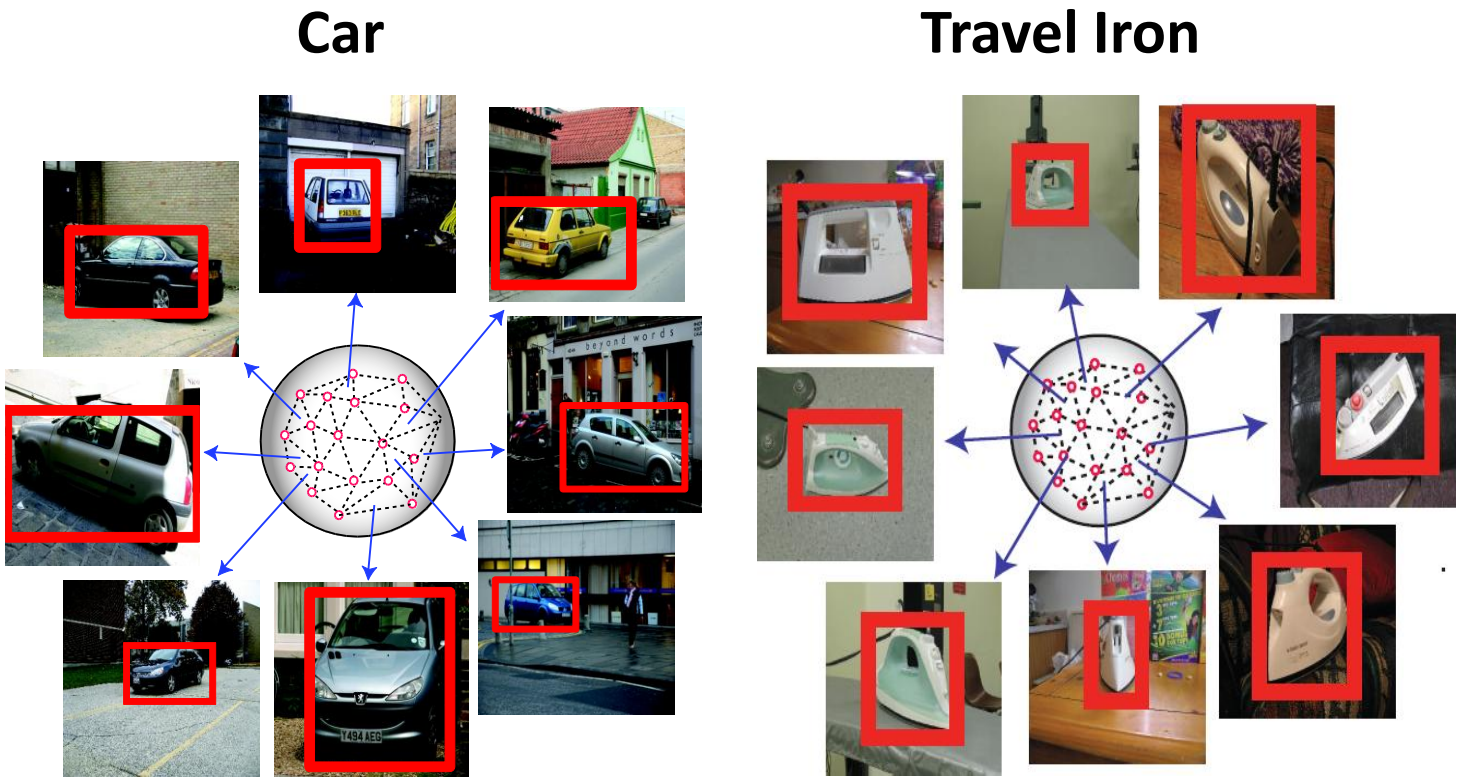
Typical Examples

Bicycle **Binocular Microscope**



Blue arrows indicate the viewpoint for the detected object (in red bounding box).

Typical Examples



Blue arrows indicate the viewpoint for the detected object (in red bounding box.

Novel view object synthesis from a single image

For the first time!

car calculatorflashlightbike

For natural or artificial scenes, see hoeim 07; saxena

Conclusions

	Single view	Mixture / Multi-view	Sav. et al, 07	Morphing model
View point invariant	X	√	√	√
No supervision	√	X Category View point	X → √ category ; all views all instances available	X → √ category
# Categories	~300	2	8	16
Share information across views	X	√	√	√
View synthesis	X	X	X	X → √
Pose estimation	X → √	X	X → √	√

Thank you!