CEE598 - Visual Sensing for Civil Infrastructure Eng. & Mgmt.

Session 20 – Object Recognition 3

Mani Golparvar-Fard
Department of Civil and Environmental Engineering
3129D, Newmark Civil Engineering Lab
e-mail: mgolpar@illinois.edu

Outline

- Object Recognition

- Introduction
- Recognition of single 3D objects
  - Bag of word models
  - Part based models
  - Models for 3D objects categorization

Segments of this lectures are courtesy of Prof A. Torralba, R. Fergus and F. Li
"Recognizing and Learning Object Categories: Year 2007"

Challenges: object intra-class variation

Usual Challenges

- Variability due to:
  - View point
  - Illumination
  - Occlusions

Problem with bag-of-words

- All have equal probability for bag-of-words methods
- Location information is important

Part Based Representation

- Object as set of parts
- Model:
  - Relative locations between parts
  - Appearance of part

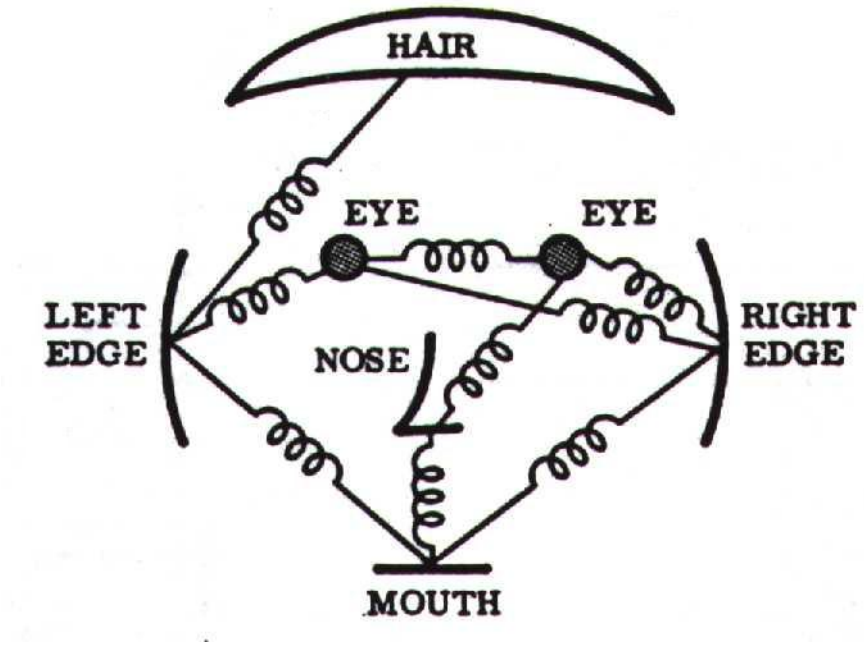
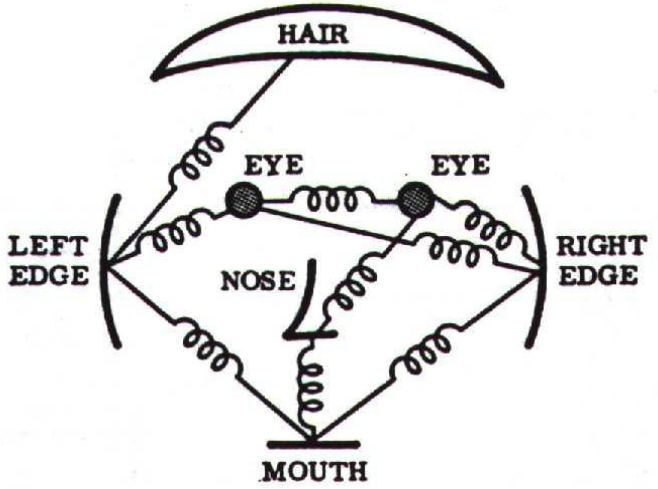


Figure from [Fischler & Elschlager 73]

History of Parts and Structure approaches

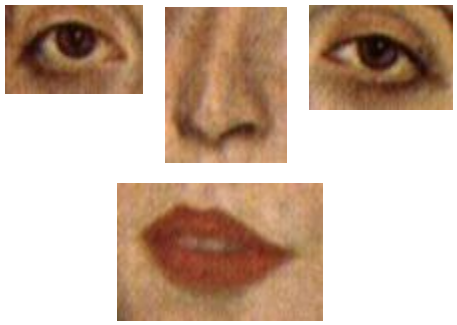
- Fischler & Elschlager 1973



- Yuille '91
- Brunelli & Poggio '93
- Lades, v.d. Malsburg et al. '93
- Cootes, Lanitis, Taylor et al. '95
- Amit & Geman '95, '99
- Perona et al. '95, '96, '98, '00, '03, '04, '05
- Ullman et al. 02
- Felzenszwalb & Huttenlocher '00, '04
- Crandall & Huttenlocher '05, '06
- Leibe & Schiele '03, '04

- Many papers since 2000

Deformations



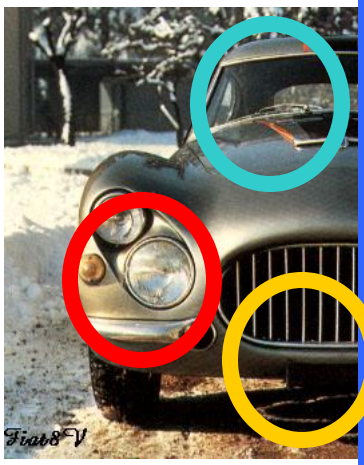
A

B

C

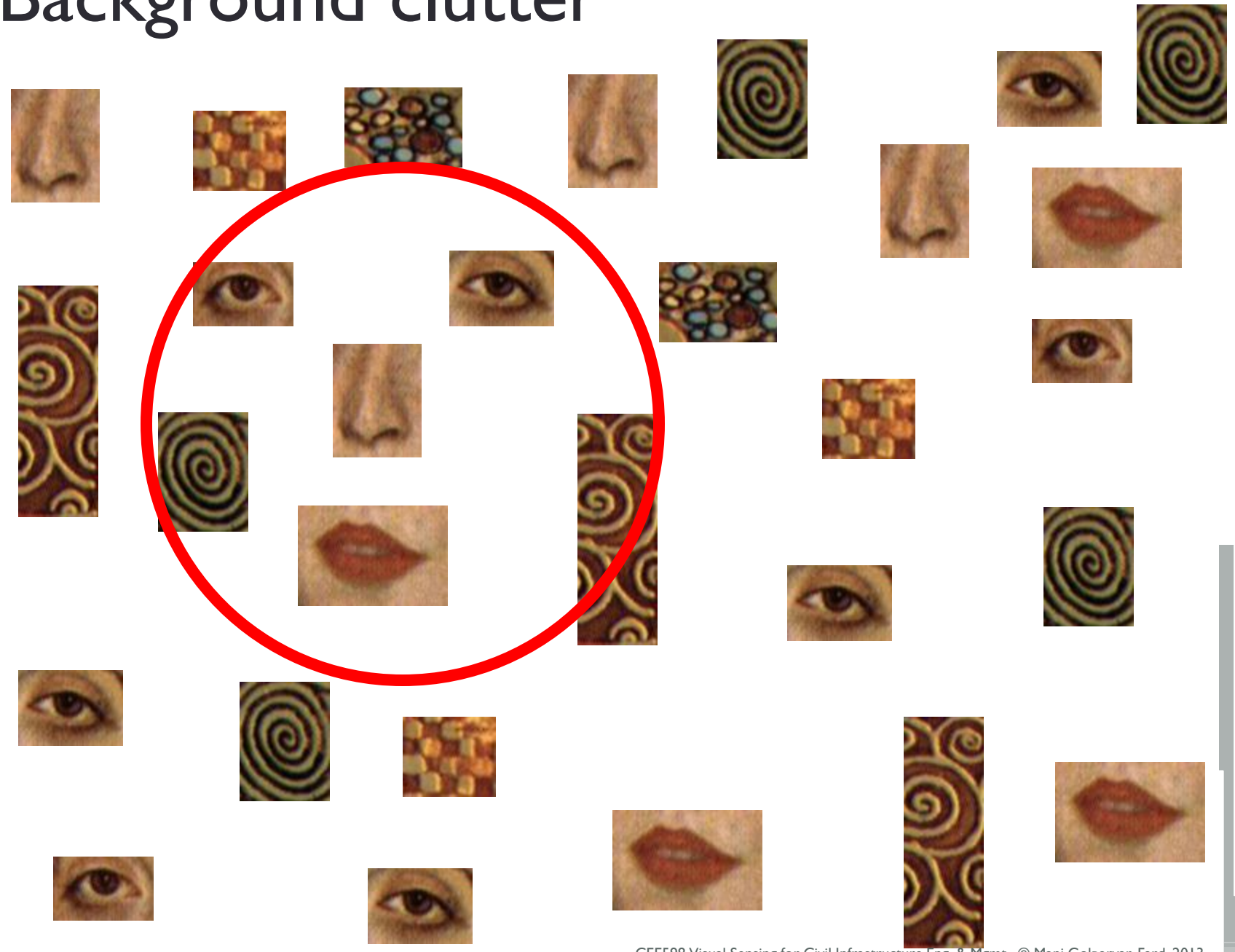
D

Presence / Absence of Features



occlusion

Background clutter



Sparse representation

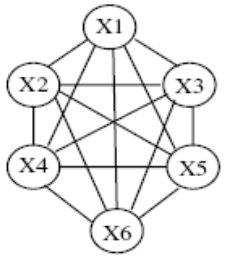
- Computationally tractable (10^5 pixels → 10^1 -- 10^2 parts)
- Generative representation of class
- Avoid modeling global variability

- Throw away most image information
- Parts need to be distinctive to separate from other classes

Different connectivity structures

Fergus et al. '03
Fei-Fei et al. '03

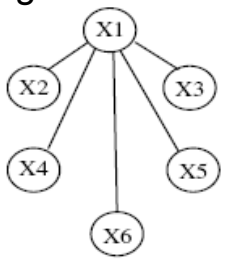
$O(N^6)$



a) Constellation [13]

Crandall et al. '05
Fergus et al. '05

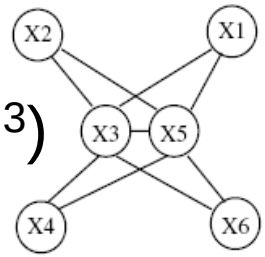
$O(N^2)$



b) Star shape [9, 14]

Crandall et al. '05

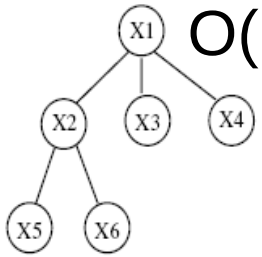
$O(N^3)$



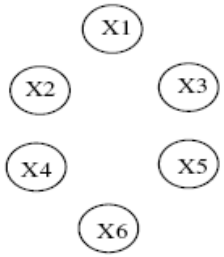
c) k-fan ($k = 2$) [9]

Felzenszwalb & Huttenlocher '00

$O(N^2)$

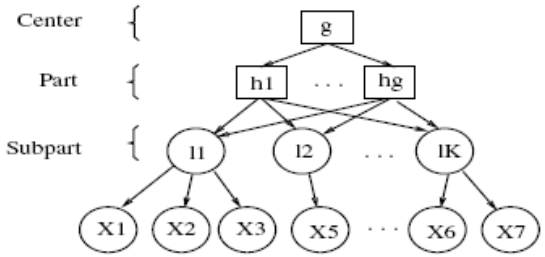


d) Tree [12]



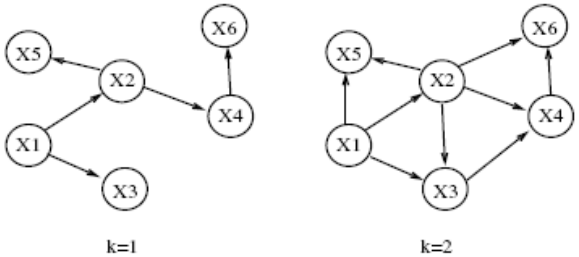
e) Bag of features [10, 21]

Csurka '04
Vasconcelos '00



f) Hierarchy [4]

Bouchard & Triggs '05



g) Sparse flexible model

Carneiro & Lowe '06

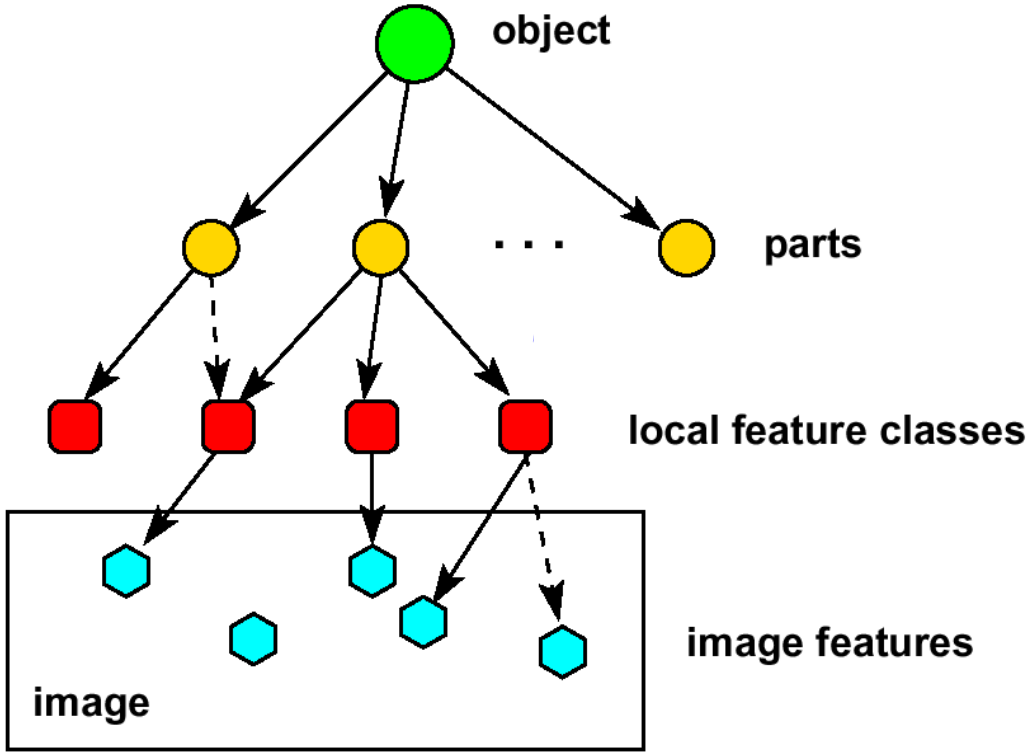
from Sparse Flexible Models of Local Features
Gustavo Carneiro and David Lowe, ECCV 2006

Hierarchical representations

- Pixels → Pixel groupings → Parts → Object

- Multi-scale approach increases number of low-level features

- Amit and Geman '98
- Bouchard & Triggs '05



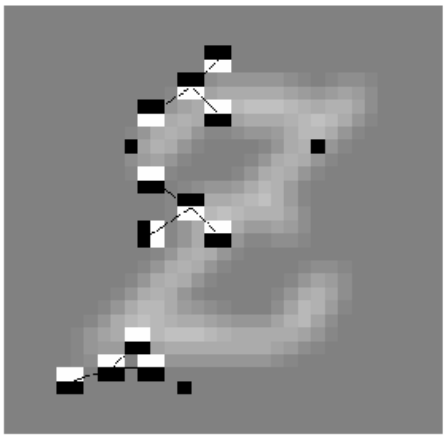
Images from [Amit98,Bouchard05]

Hierarchical representations

- Pixels → Pixel groupings → Parts → Object

- Multi-scale approach increases number of low-level features

- Amit and Geman '98
- Bouchard & Triggs '05



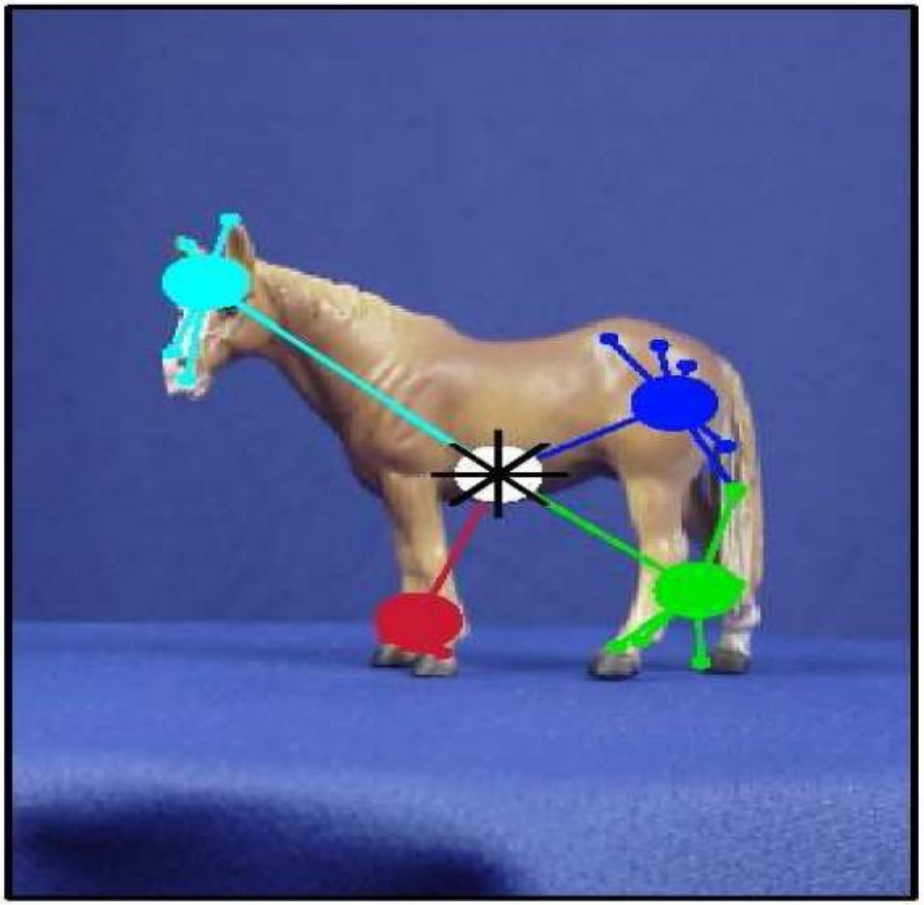
Images from [Amit98,Bouchard05]

Hierarchical representations

- Pixels → Pixel groupings → Parts → Object

- Multi-scale approach increases number of low-level features

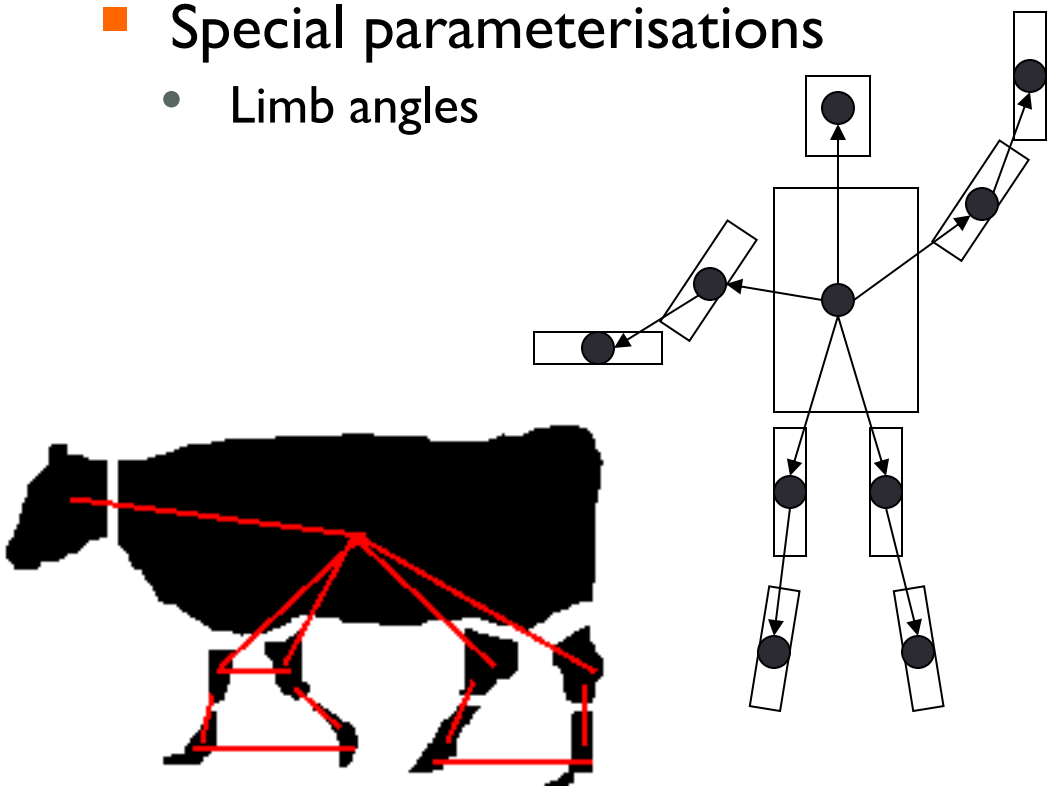
- Amit and Geman '98
- Bouchard & Triggs '05

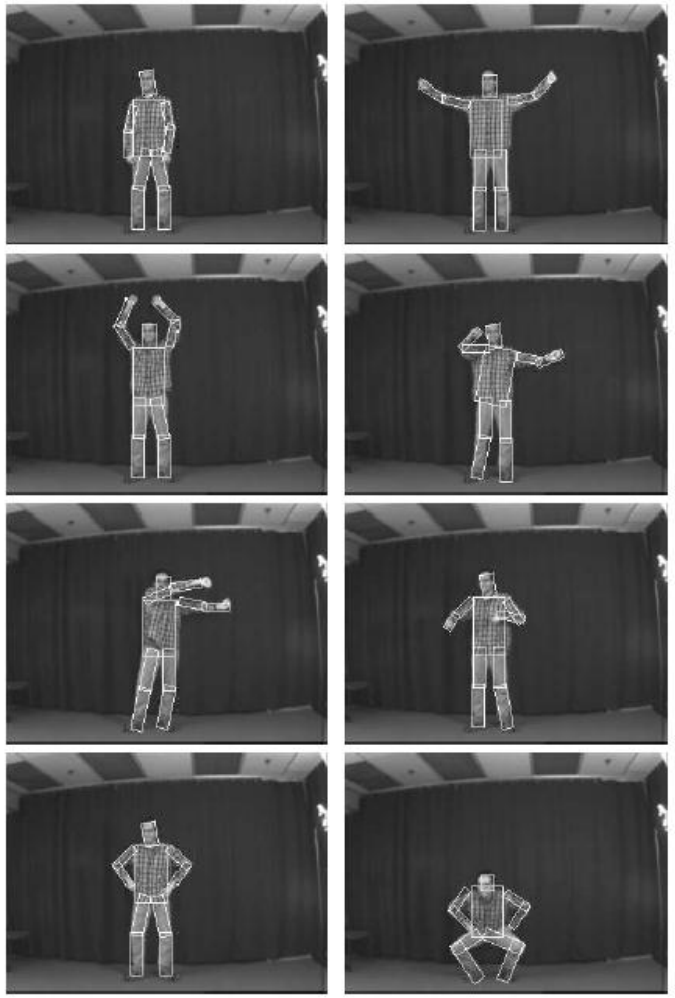


Images from [Amit98,Bouchard05]

Some class-specific graphs

- Articulated motion
  - People
  - Animals
- Special parameterisations
  - Limb angles

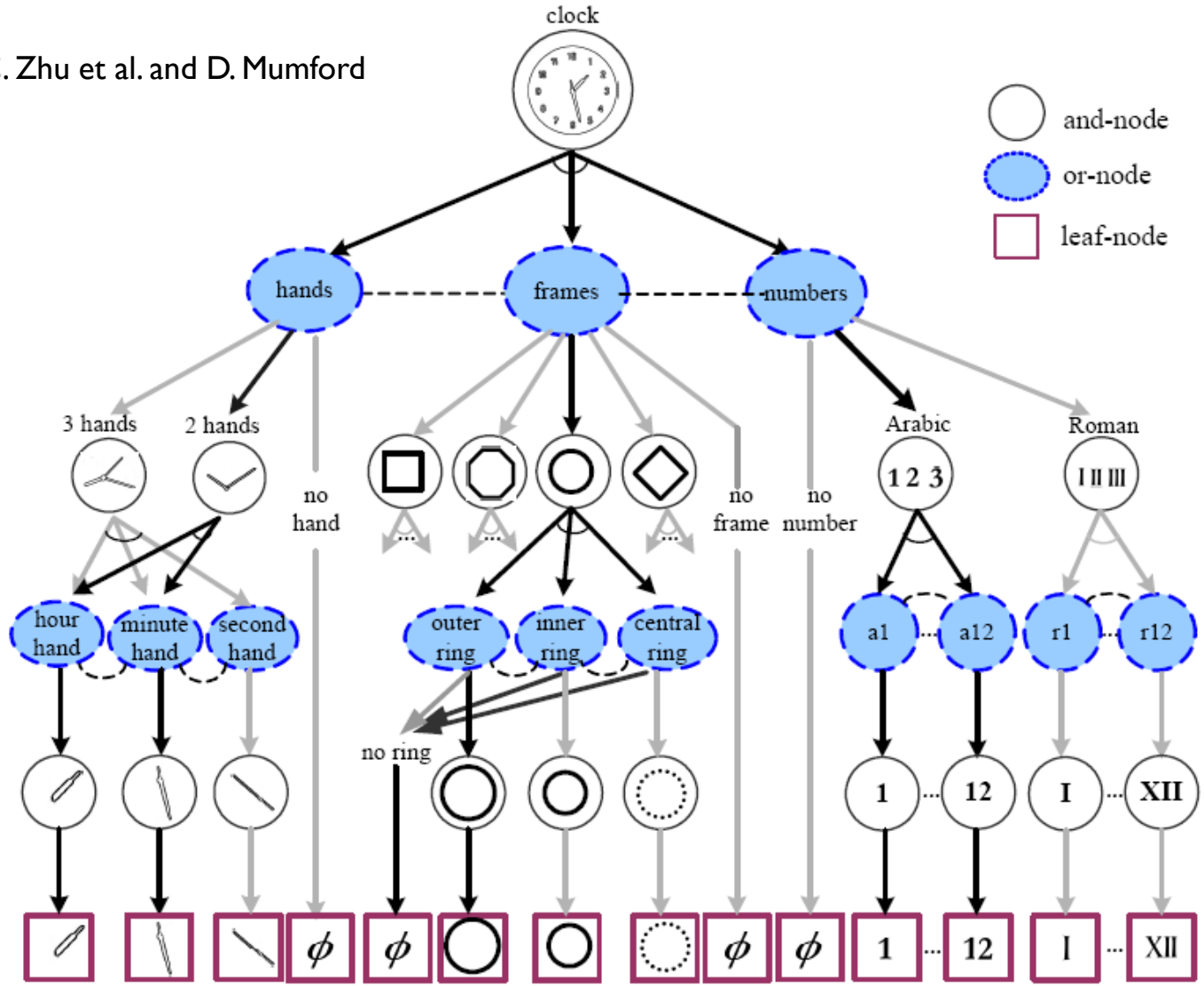




Images from [Kumar, Torr and Zisserman 05, Felzenszwalb & Huttenlocher 05]

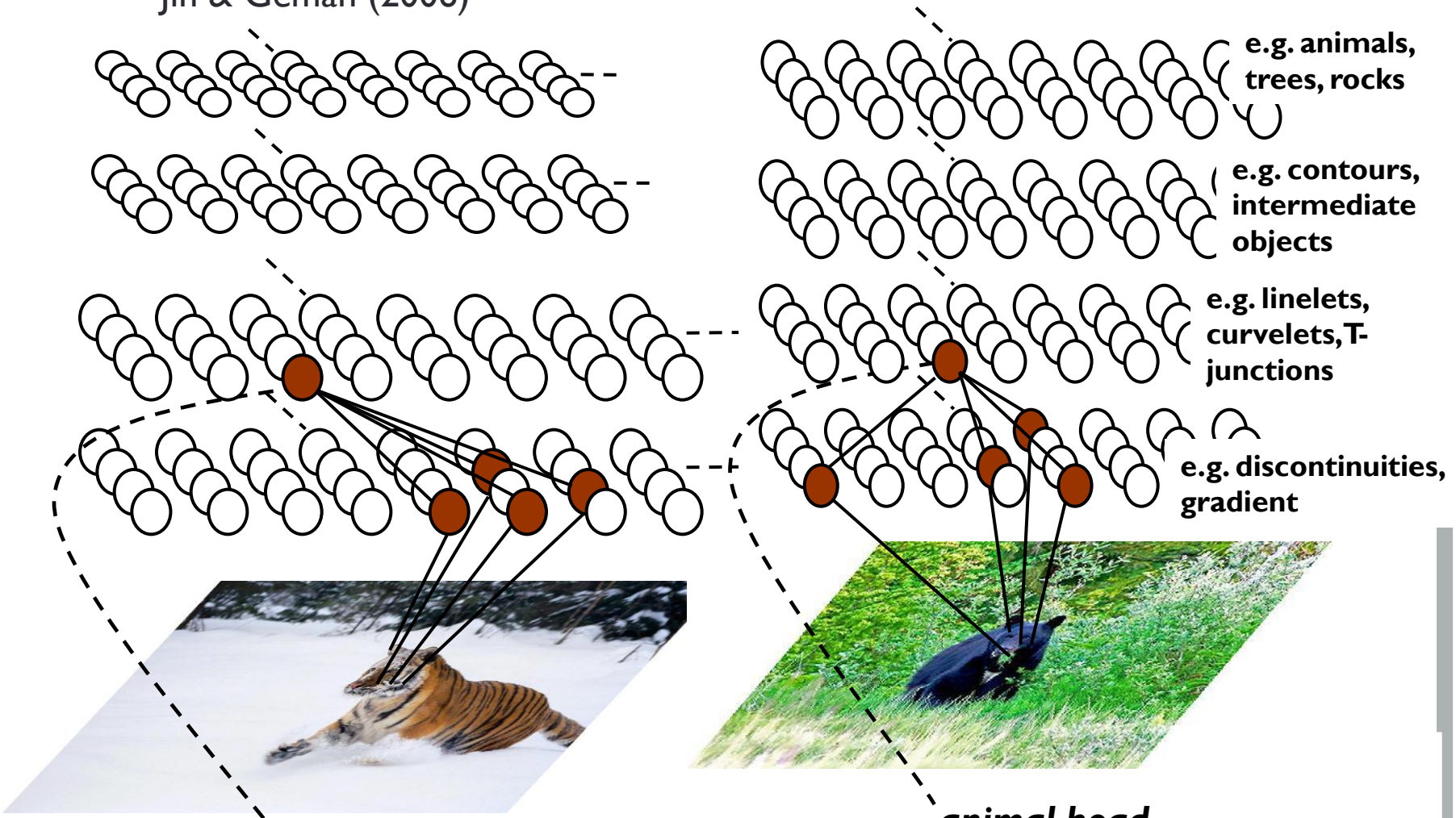
Stochastic Grammar of Images

- S.C. Zhu et al. and D. Mumford



Context and Hierarchy in a Probabilistic Image Model

- Jin & Geman (2006)



Two approaches

- Generative part-based models (constellation models)

- Implicit shape models

Generative part-based models

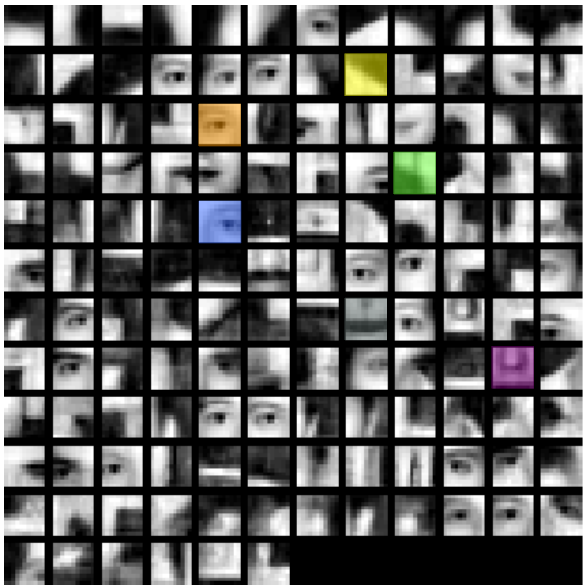
- E.g. Gaussian distribution (parameters of model, μ and Σ)

- Burl et al. '96
- M. Weber, M. Welling P. Perona, '00
- R. Fergus, P. Perona and A. Zisserman, '03

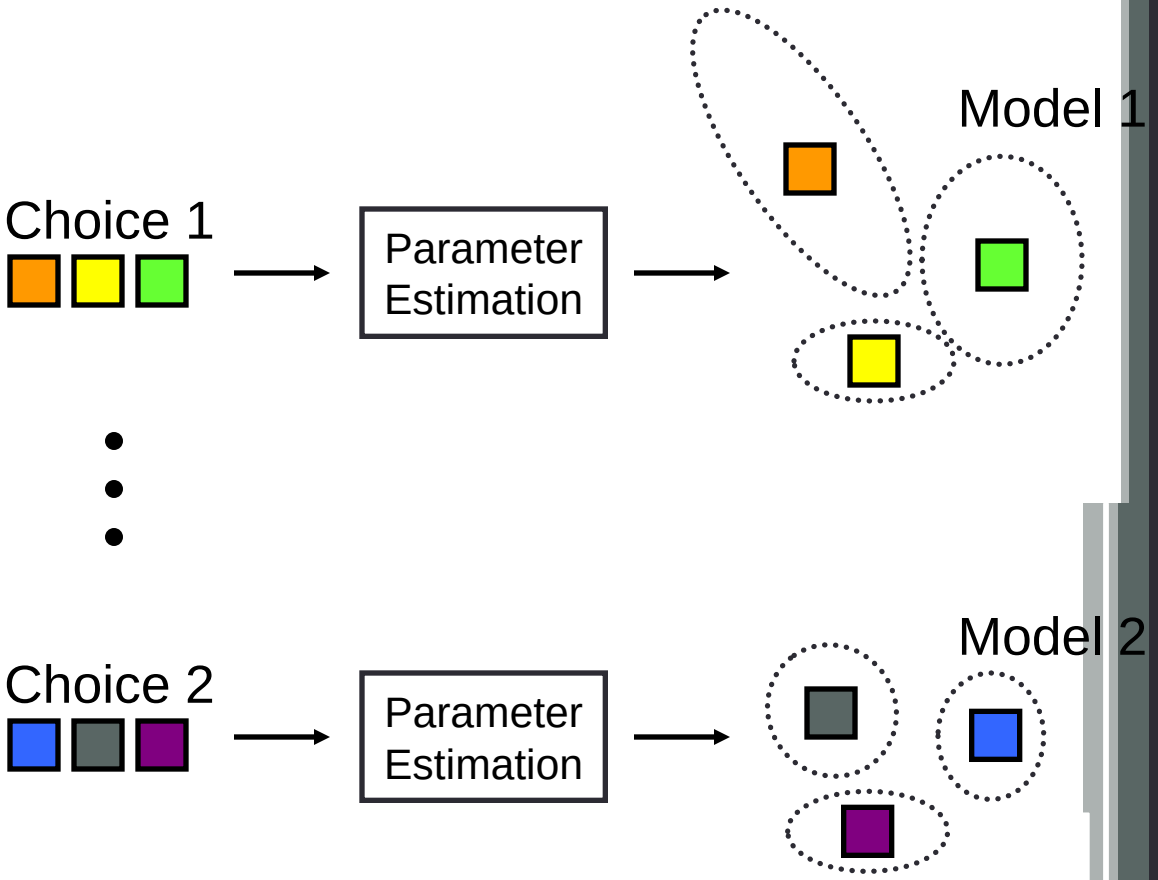
Learn appearance & shape

[appearance first, then shape]

Weber et al. ‘00



Preselected Parts (≈100)
[Code book]



Recall: Object categorization: the statistical viewpoint

- **Discriminative methods** model posterior
- **Generative methods** model likelihood and prior

- Bayes rule:

$$\underbrace{\frac{p(excavator \mid image)}{p(no\ excavator \mid image)}}_{\text{posterior ratio}} = \underbrace{\frac{p(image \mid excavator)}{p(image \mid no\ excavator)}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(excavator)}{p(no\ excavator)}}_{\text{prior ratio}}$$

Probabilistic model

$$P(\text{image} \mid \text{object}) = P(\text{appearance}, \text{shape} \mid \text{object})$$

Part descriptors

Part locations

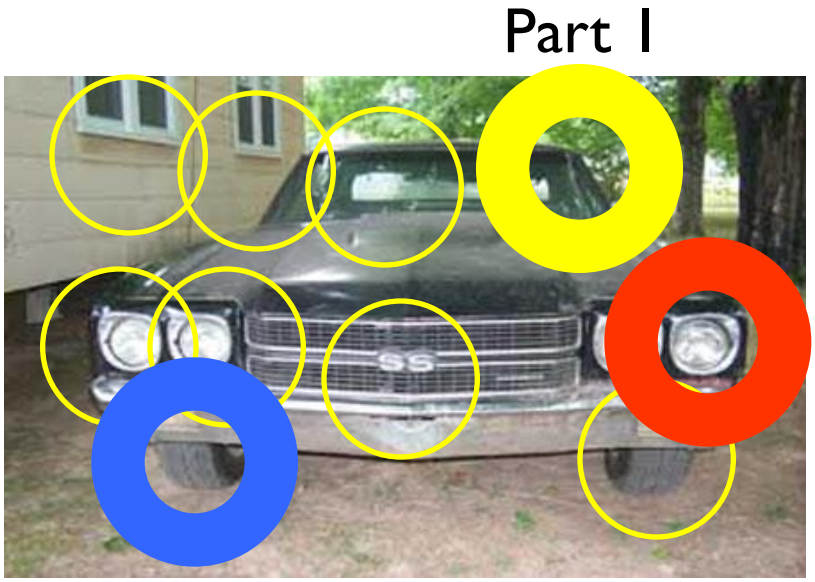
Candidate parts (features)

Credit slide: S. Lazebnik

Probabilistic model

$$P(\text{image} \mid \text{object}) = P(\text{appearance}, \text{shape} \mid \text{object})$$

Introduce variable h to regulate feature assignment



Credit slide: S. Lazebnik

Probabilistic model

$$P(\text{image} \mid \text{object}) = P(\text{appearance}, \text{shape} \mid \text{object})$$

$$P(\text{appearance} \mid h, \text{object}) p(\text{shape} \mid h, \text{object}) p(h \mid \text{object})$$

h: assignment of features to parts

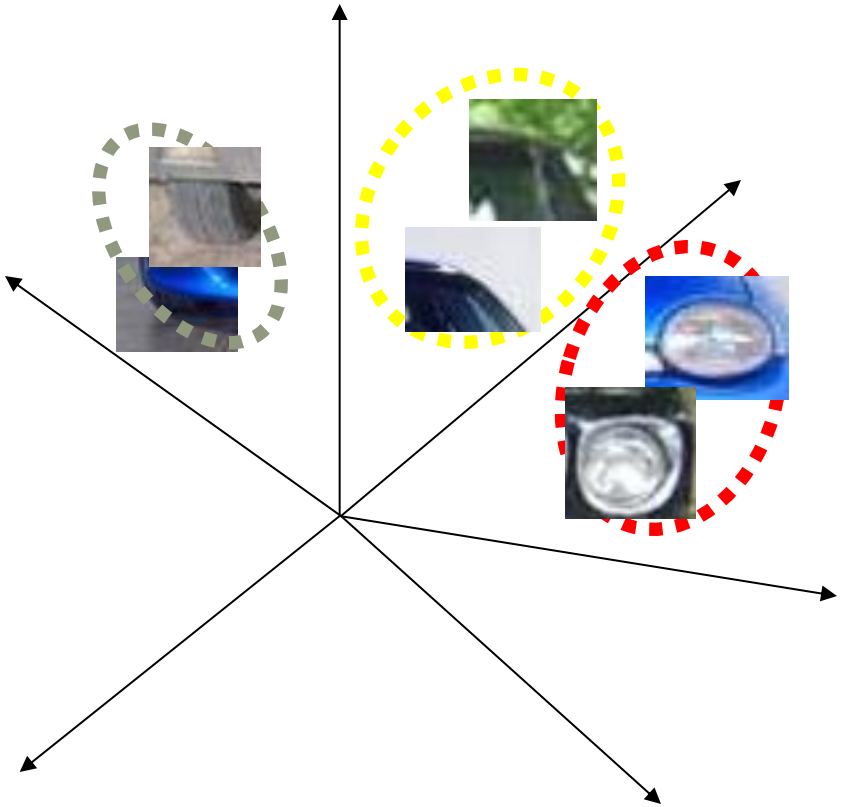
Introduce variable h to regulate feature assignment

Credit slide: S. Lazebnik

Probabilistic model

$$P(\text{image} \mid \text{object}) = P(\text{appearance}, \text{shape} \mid \text{object})$$

$$P(\text{appearance} \mid h, \text{object})\, p(\text{shape} \mid h, \text{object})\, p(h \mid \text{object})$$



Distribution over patch descriptors

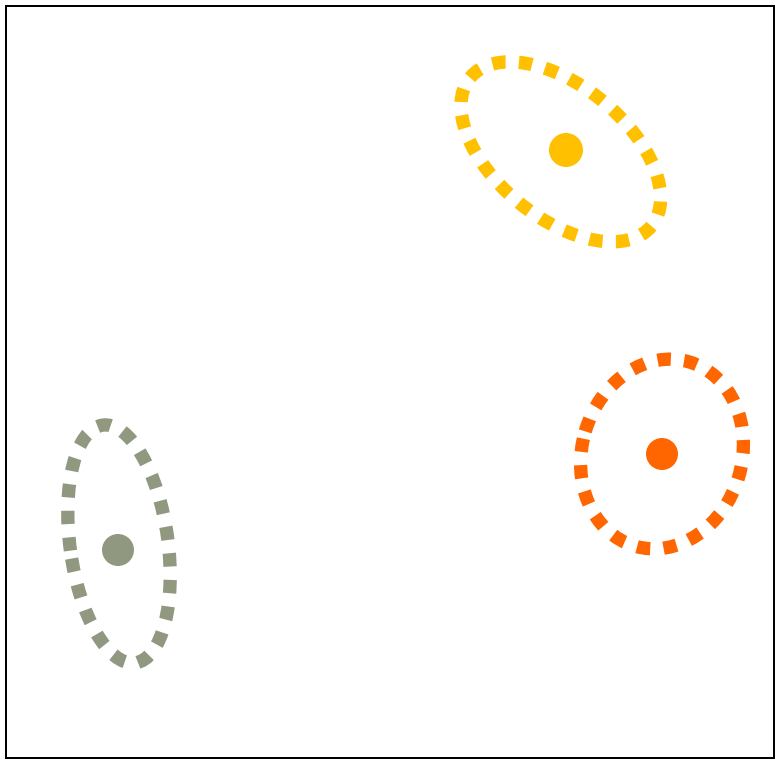
High-dimensional appearance (feature) space

Credit slide: S. Lazebnik

Probabilistic model

$$P(\text{image} \mid \text{object}) = P(\text{appearance}, \text{shape} \mid \text{object})$$

$$P(\text{appearance} \mid h, \text{object})\, p(\text{shape} \mid h, \text{object})\, p(h \mid \text{object})$$



Distribution over joint part positions

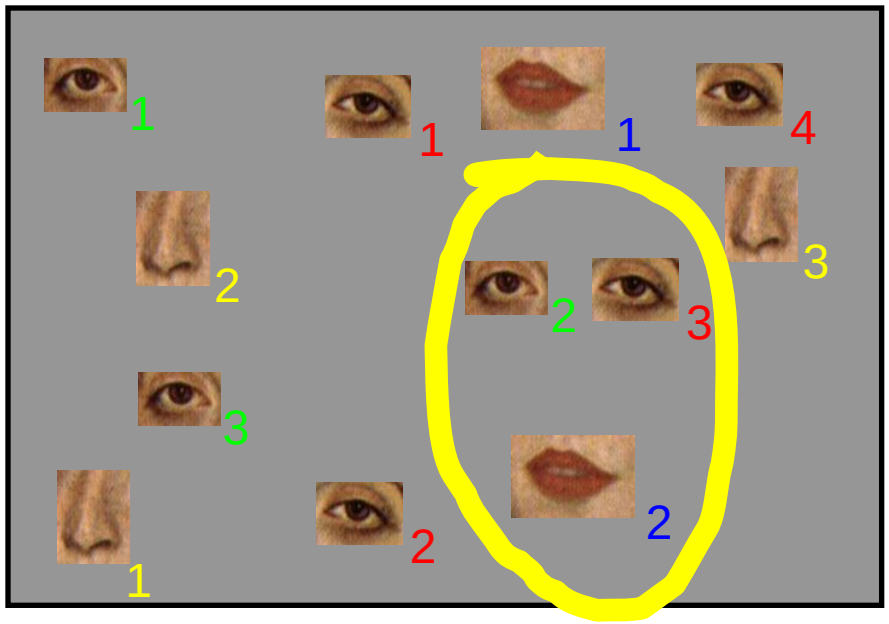
μ and Σ

2D image space

Credit slide: S. Lazebnik

Recognition

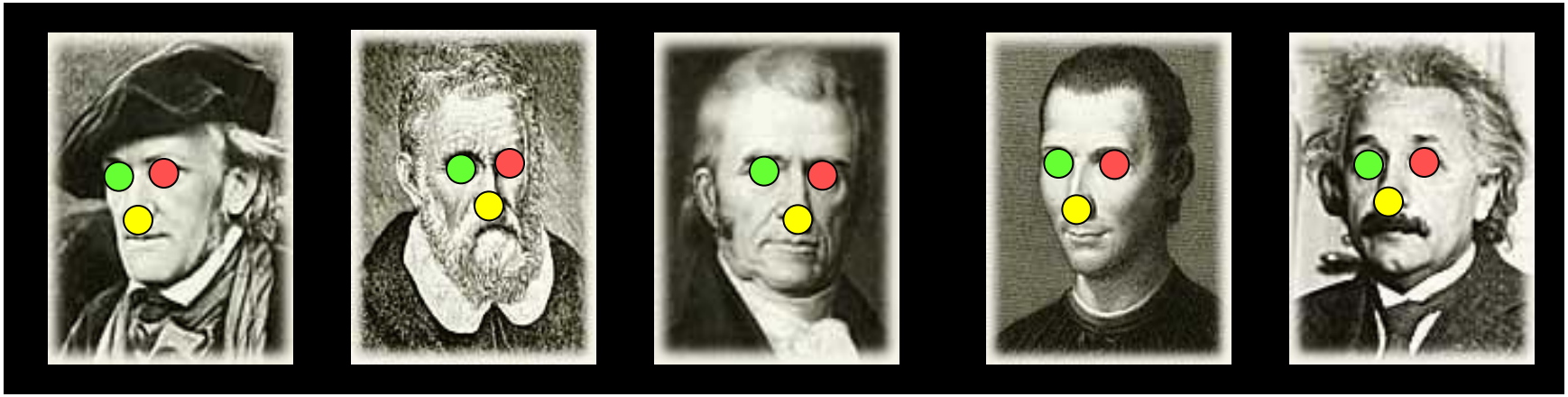
1. Run part detectors exhaustively over image



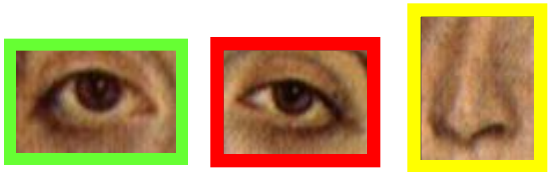
$$h = \begin{pmatrix} 0 \ldots N_1 \\ 0 \ldots N_2 \\ 0 \ldots N_3 \\ 0 \ldots N_4 \end{pmatrix} \quad \text{e.g.} \quad h = \begin{pmatrix} 2 \\ 3 \\ 0 \\ 2 \end{pmatrix}$$

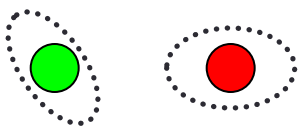
2. Try different combinations of detections in model
   - Allow detections to be missing (occlusion)
3. Pick hypothesis which maximizes: $\frac{p(Data \mid Object, Hyp)}{p(Data \mid Clutter, Hyp)}$
4. If ratio is above threshold then, instance detected

Learning Models `Manually'



- Obtain set of training images
- Choose parts
- Label parts by hand, train detectors
- Learn model from labeled parts

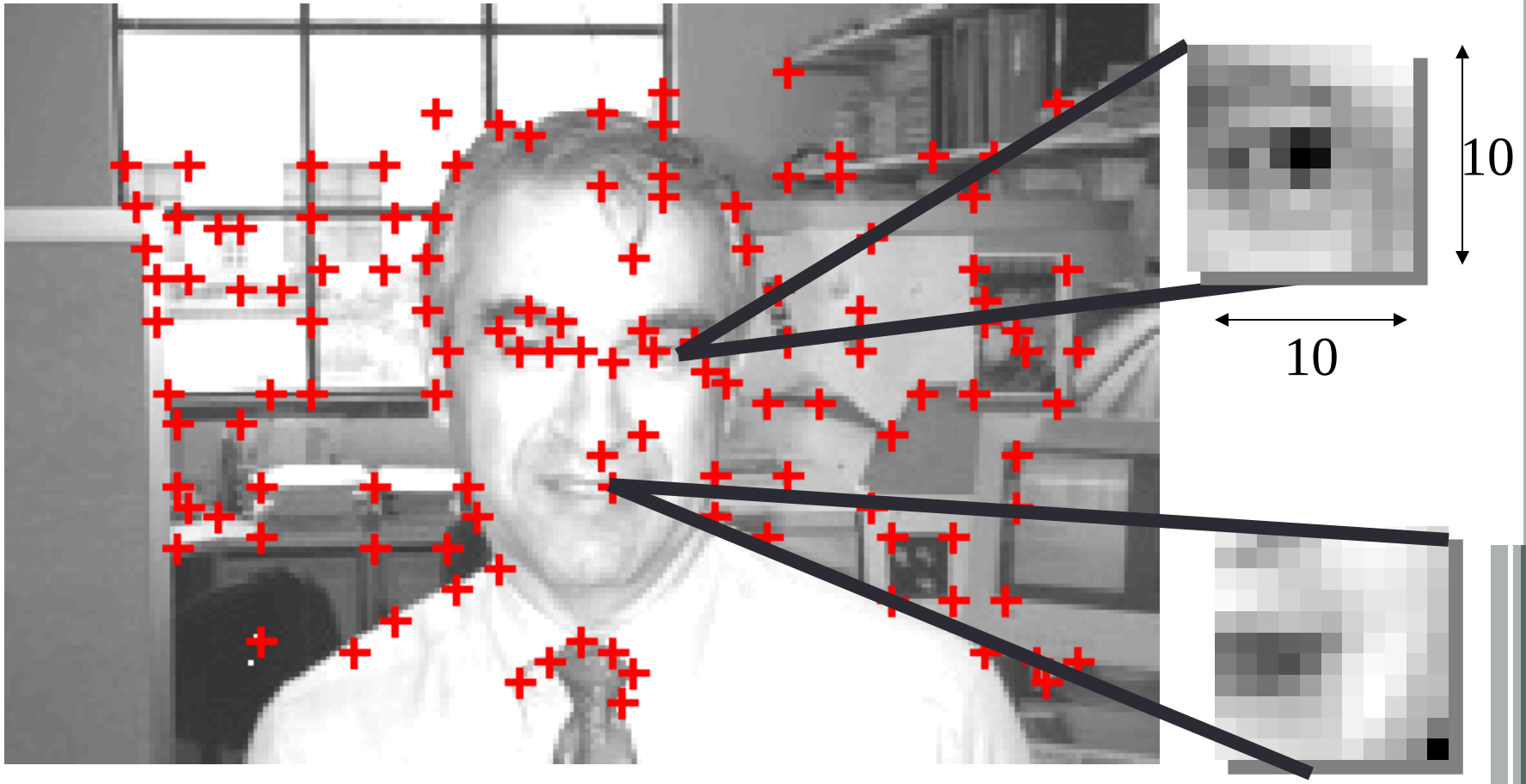




(Semi) Unsupervised learning

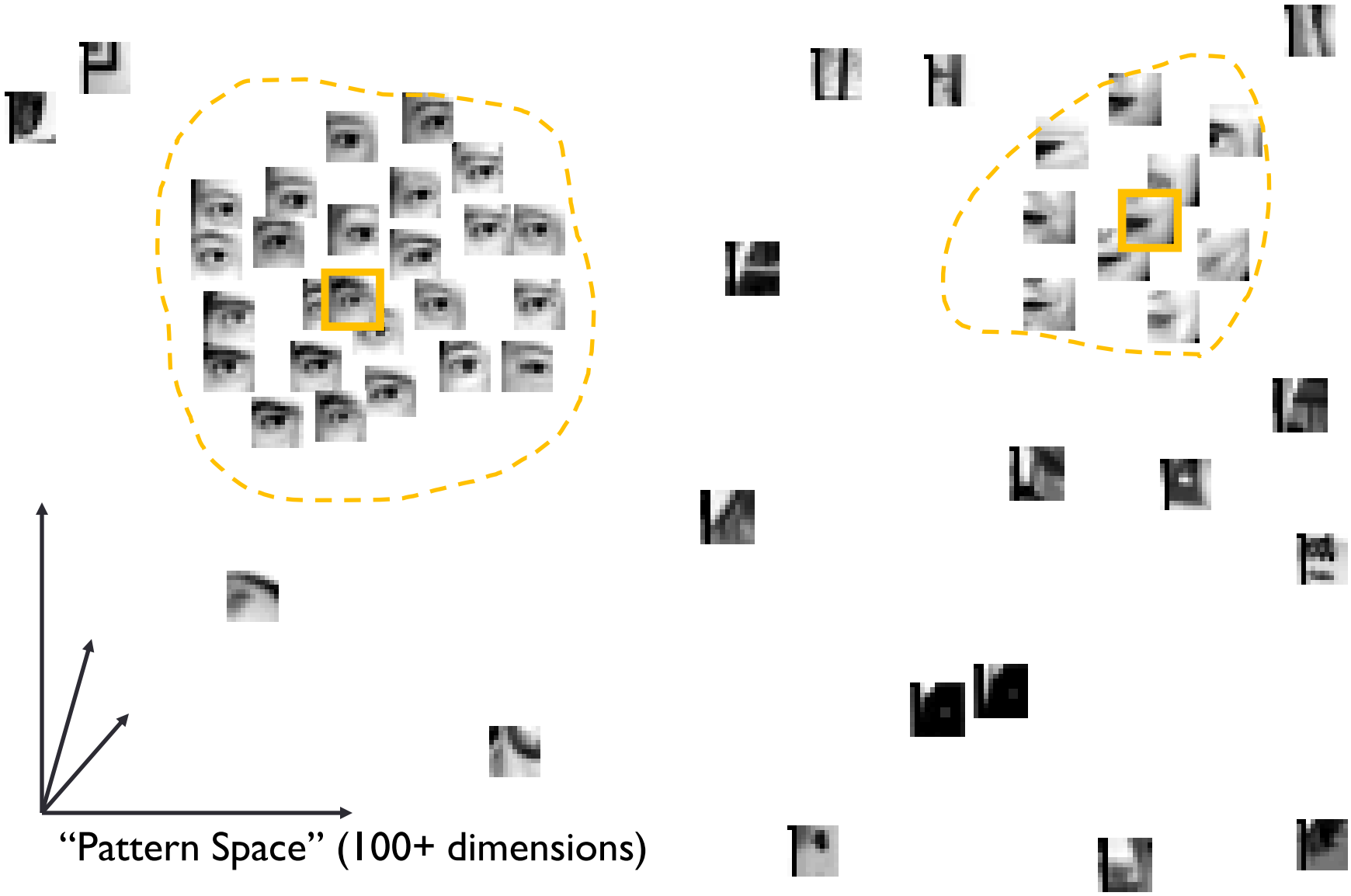
- Know if image contains object or not
- But no segmentation of object or manual selection of features

Appearance Learning procedure



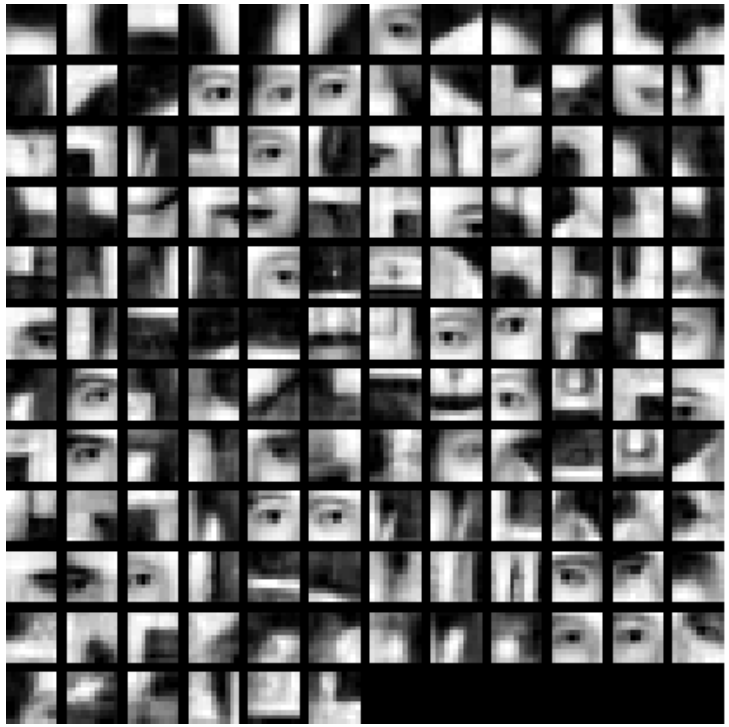
- Highly textured neighborhoods are selected automatically
- produces 100-1000 patterns per image

Appearance Learning procedure



Appearance Learning procedure

100-1000 images



~100 detectors

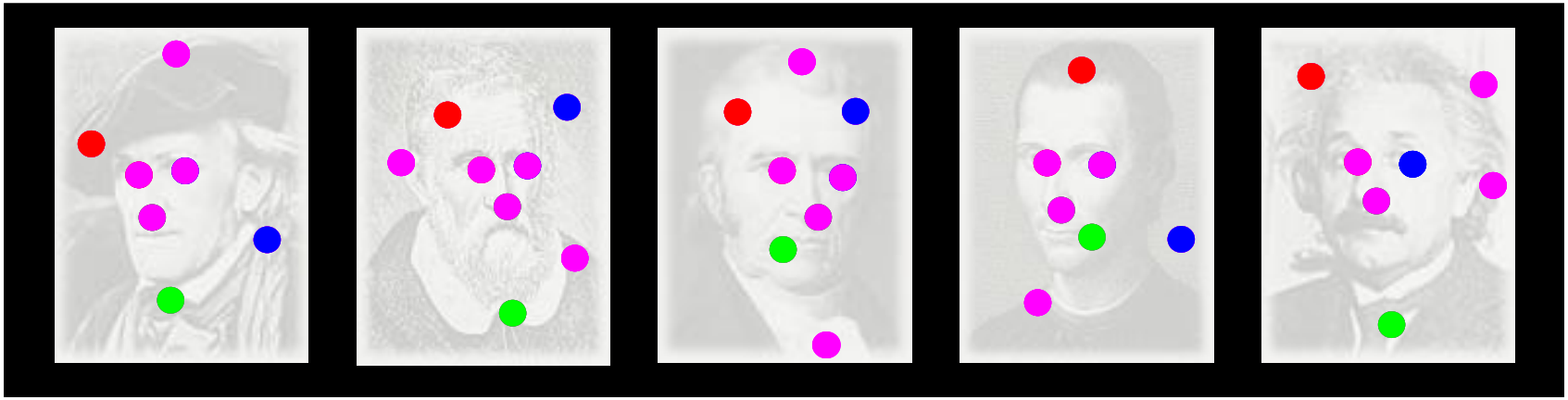
Shape Learning procedure

- Find regions & their location & appearance
- Initialize model parameters $\theta = [\mu, \Sigma]$
- Use EM and iterate to convergence:

  E-step: Use current θ to compute assignments

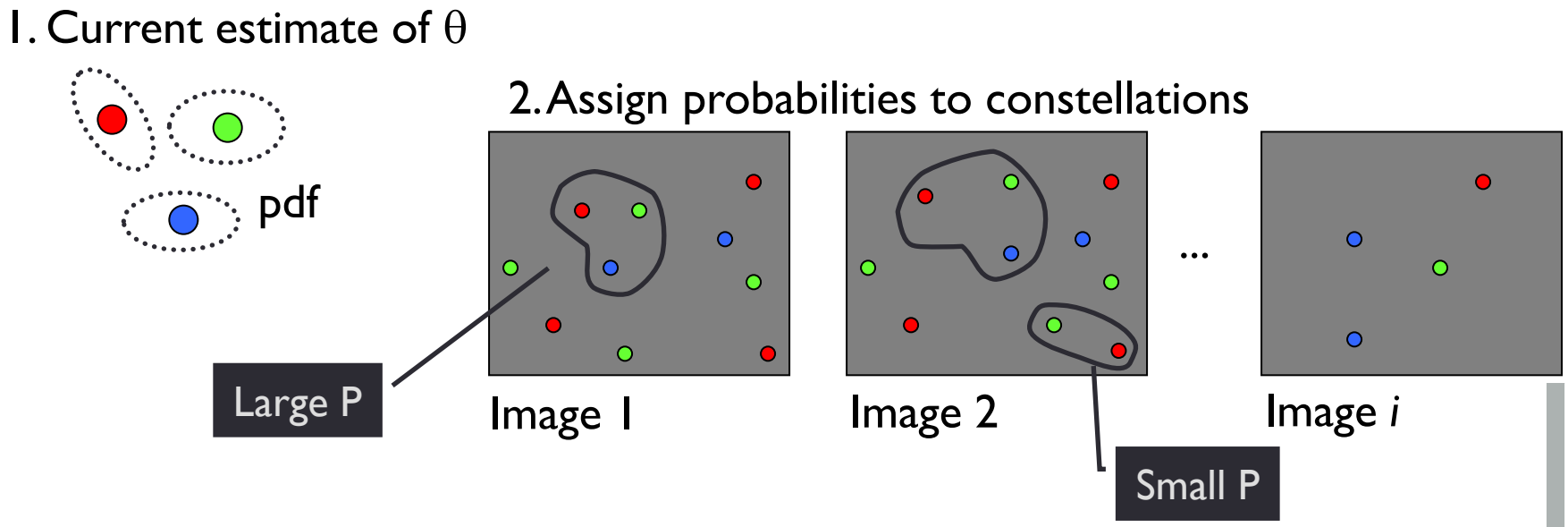
  M-step: Given assignments, update model parameters $\theta = [\mu, \Sigma]$

- Trying to maximize likelihood – consistency in shape & appearance



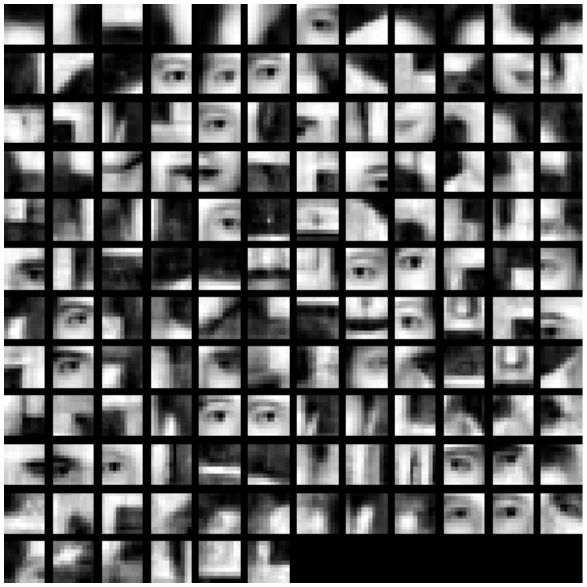
Example scheme

- using EM for maximum likelihood learning



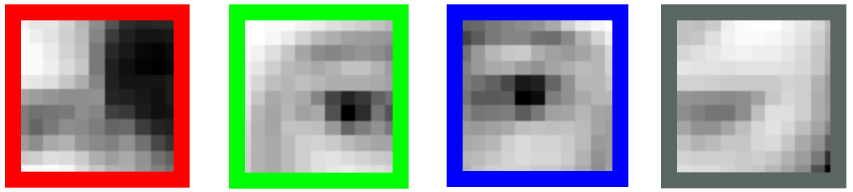
Learned Face models

Pre-selected Parts

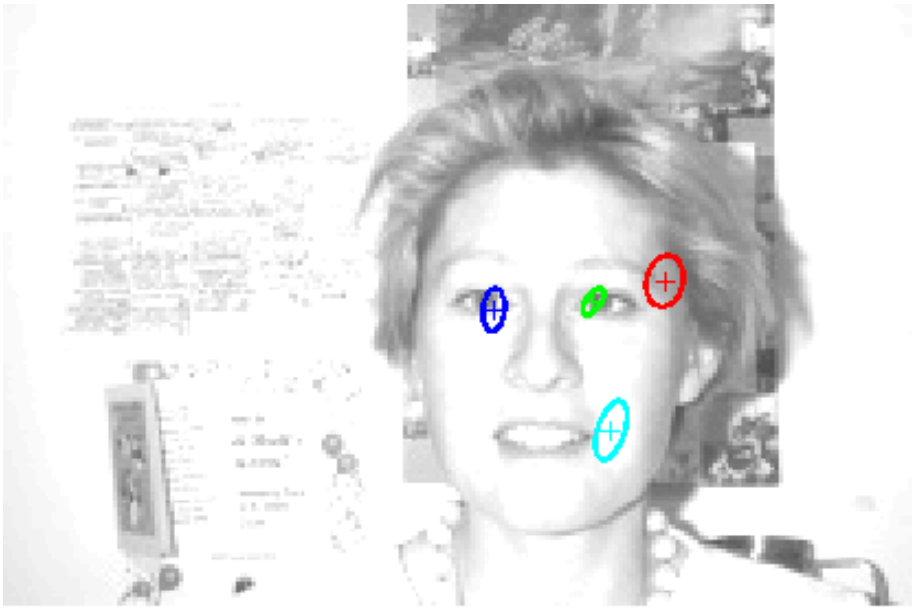


Test Error: 6% (4 Parts)

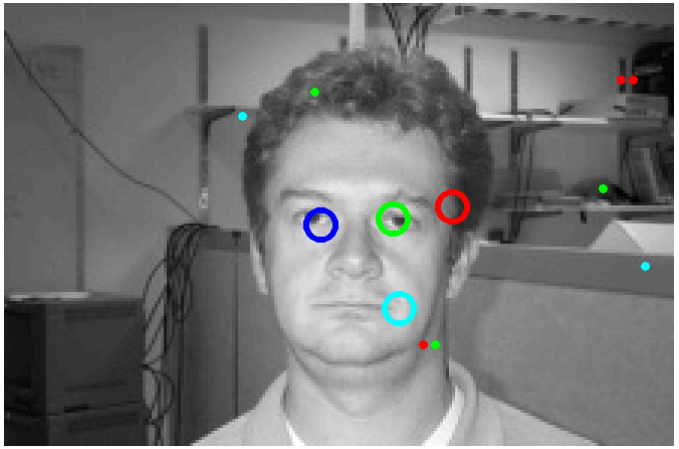
Parts in Model



Model Foreground pdf



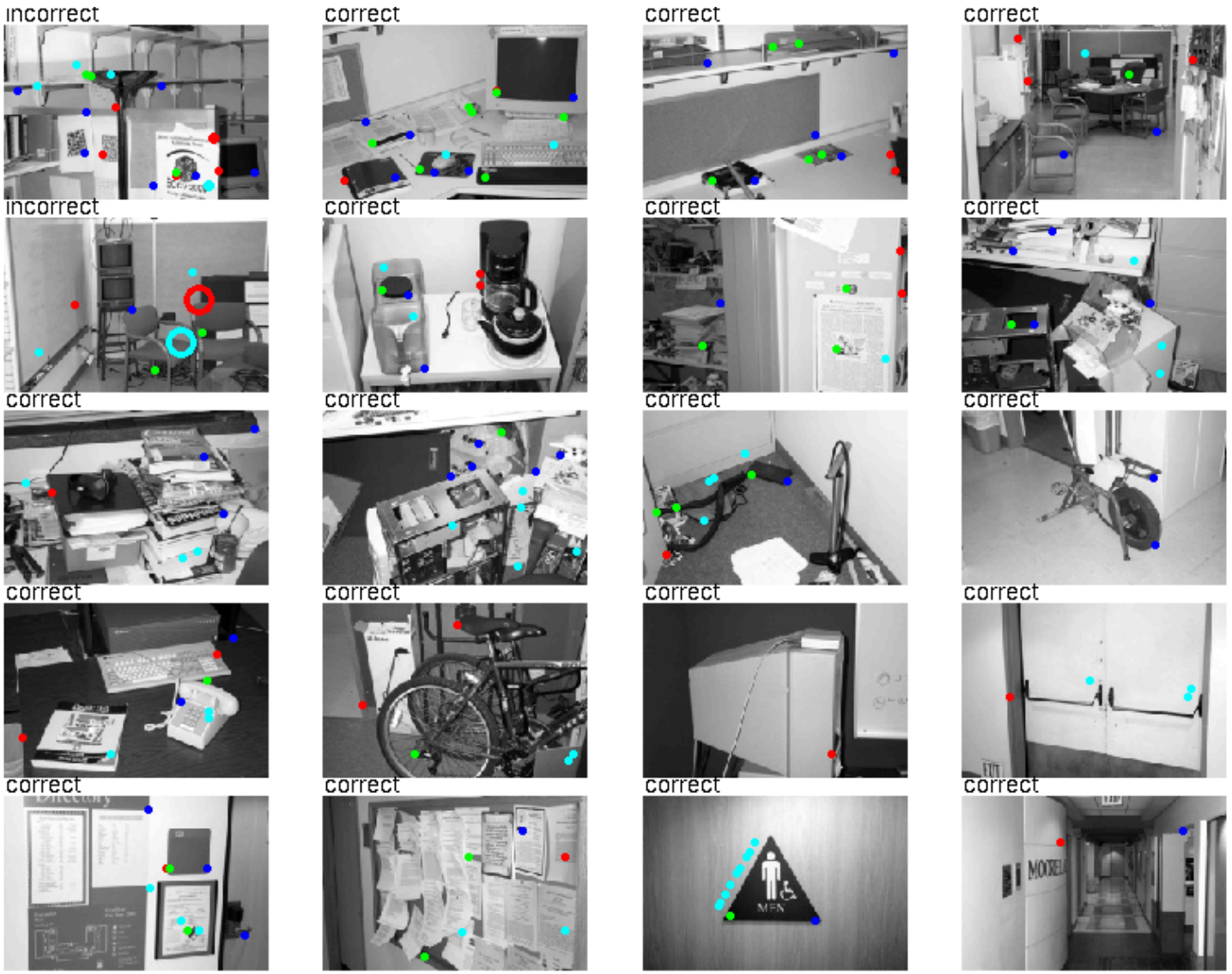
Sample Detection



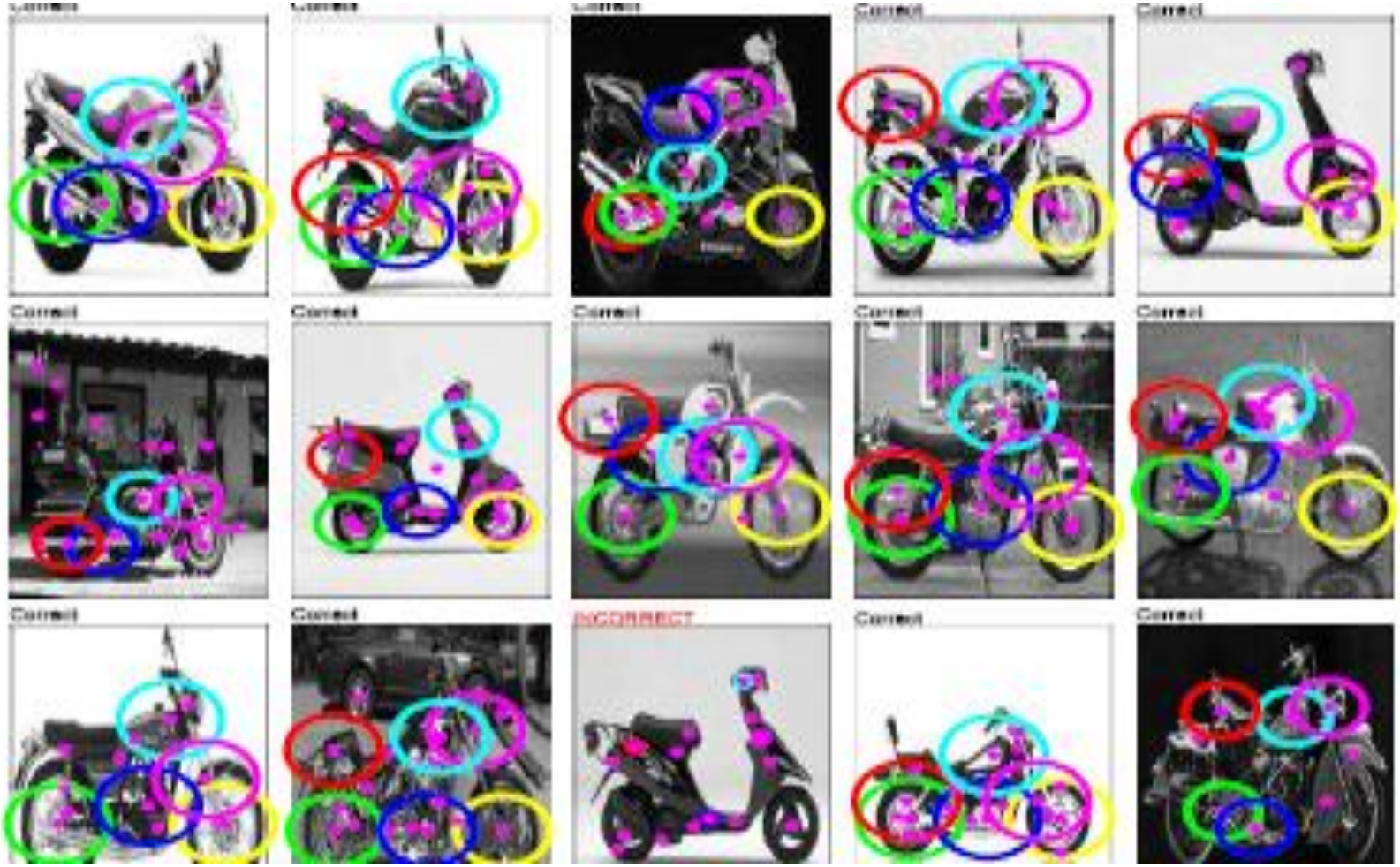
Face images


correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
incorrect
incorrect
incorrect

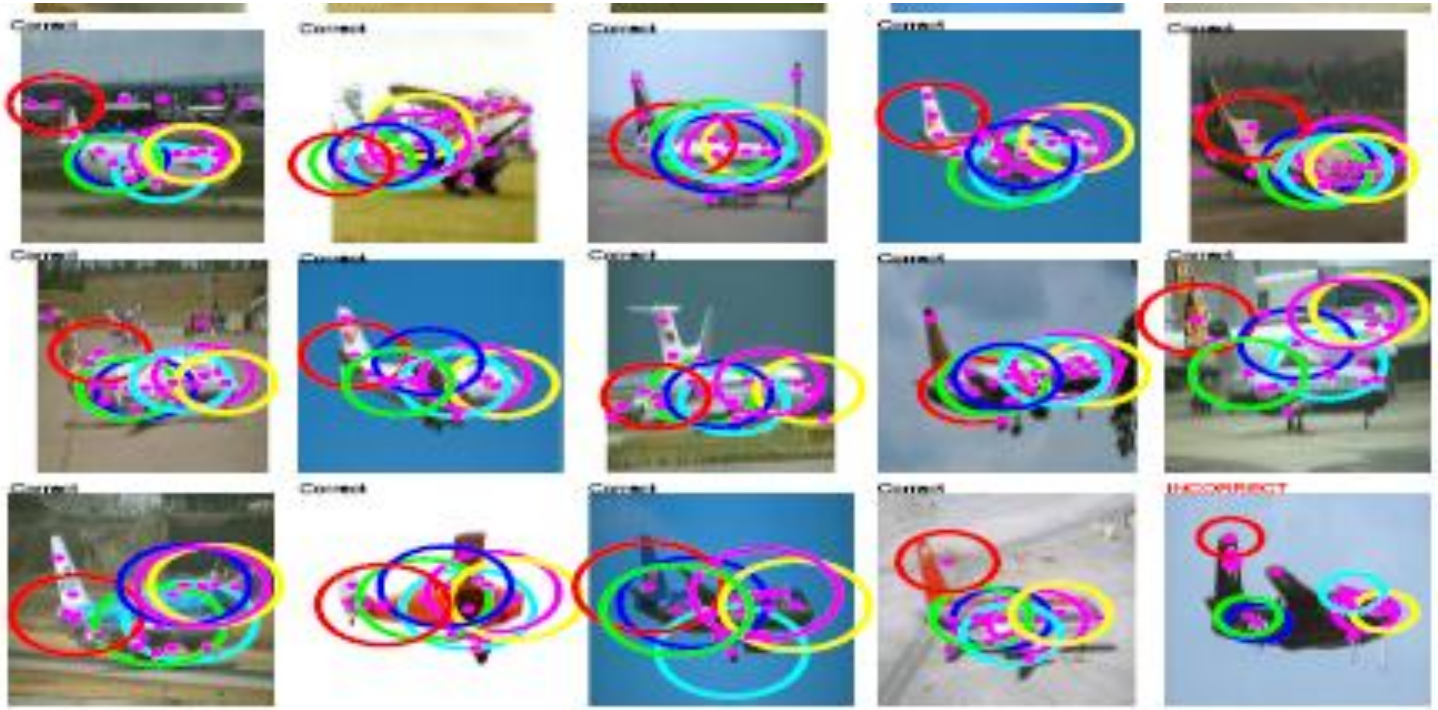
Background Images



Results: Motorbikes and airplanes



Results: Motorbikes and airplanes



Two approaches

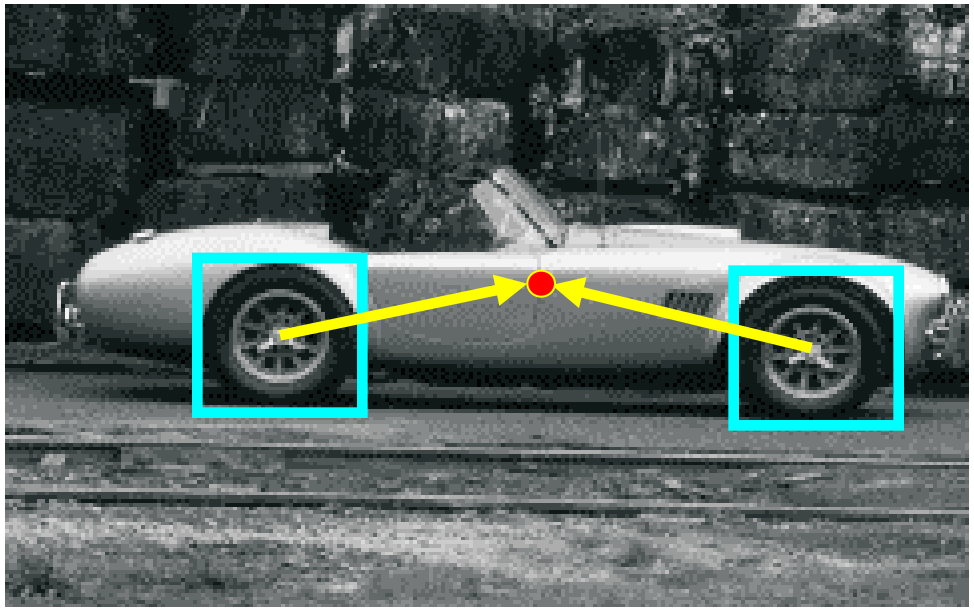
- Generative part-based models

(constellation models)
(implicit shape models)

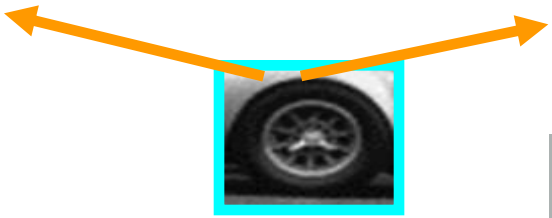
- Implicit shape models

Implicit shape models

- Visual codebook is used to index votes for object position



training image



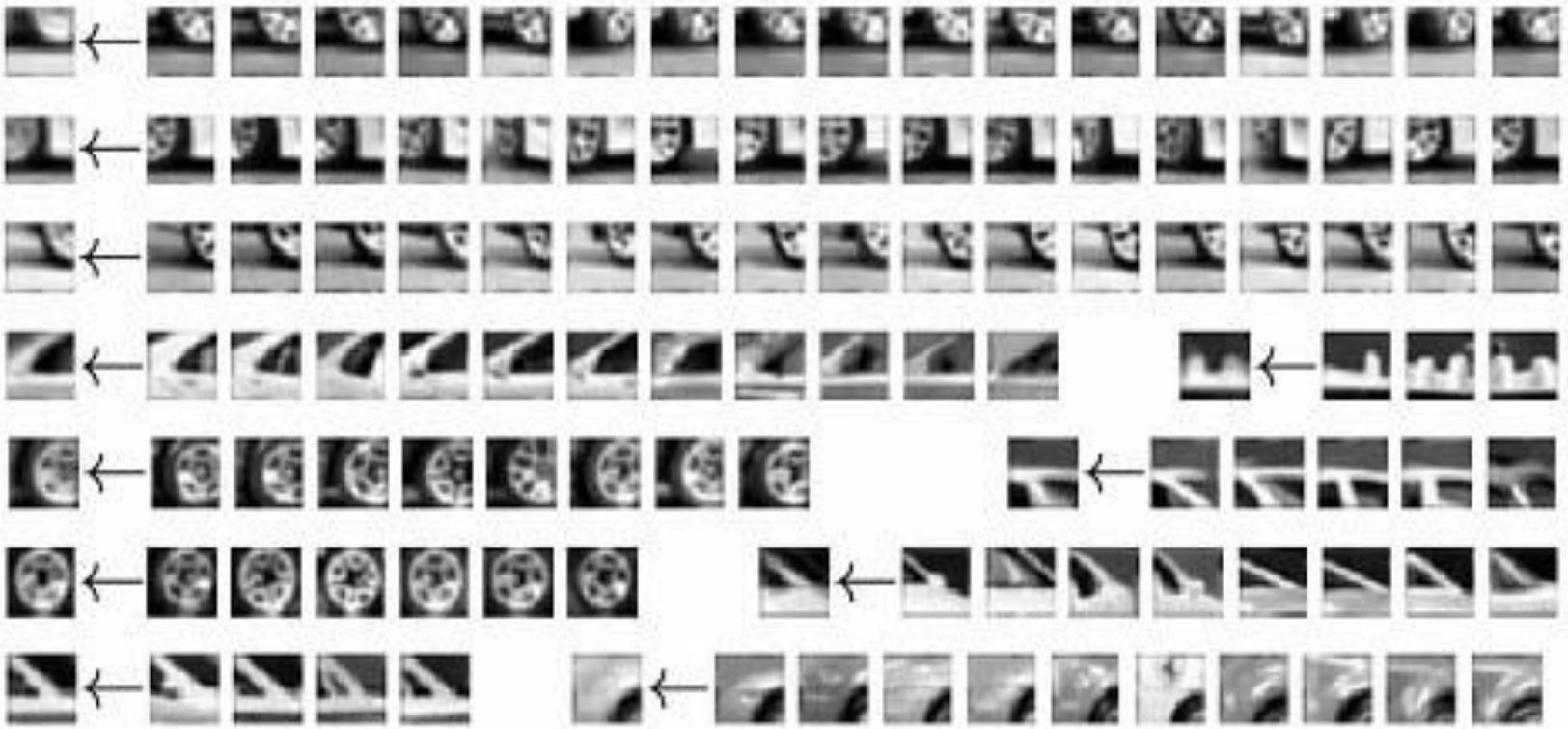
visual codeword with displacement vectors

B. Leibe, A. Leonardis, and B. Schiele, Combined Object Categorization and Segmentation with an Implicit Shape Model, ECCV Workshop on Statistical Learning in Computer Vision 2004

Credit slide: S. Lazebnik

Implicit shape models: Training

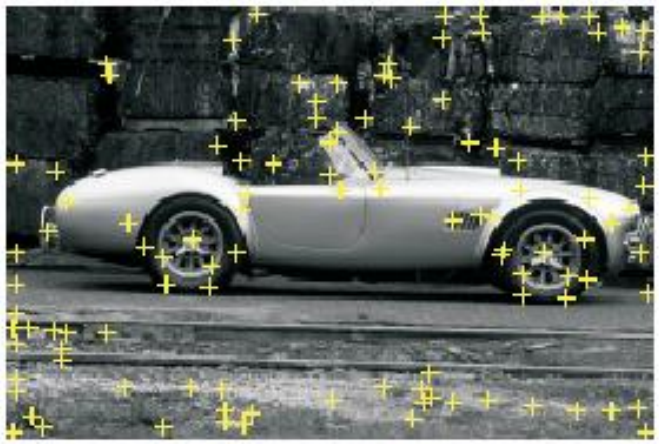
1. Build codebook of patches around extracted interest points using clustering

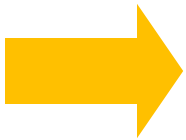


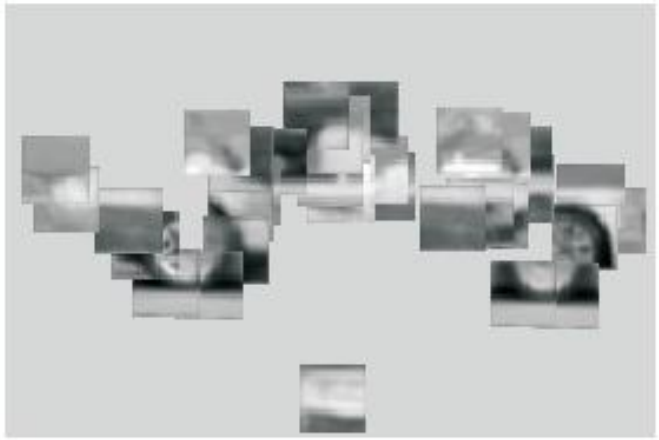
Credit slide: S. Lazebnik

Implicit shape models: Training

1. Build codebook of patches around extracted interest points using clustering
2. Map the patch around each interest point to closest codebook entry



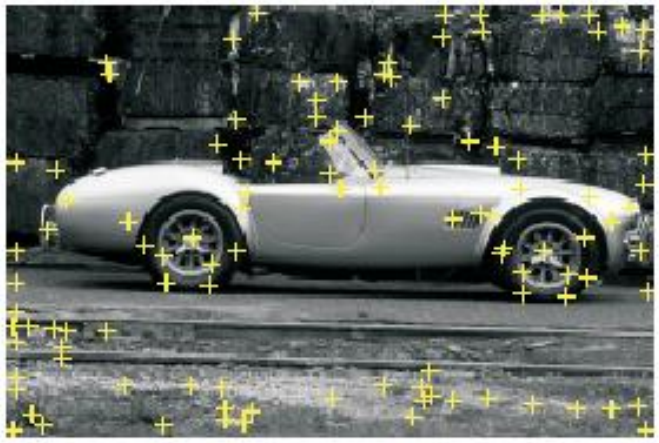


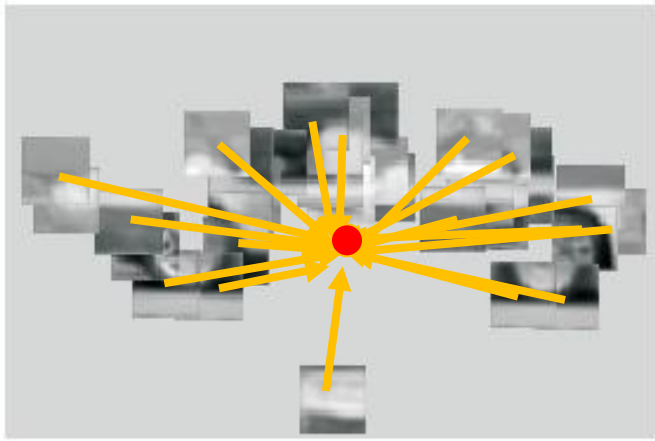


Credit slide: S. Lazebnik

Implicit shape models: Training

1. Build codebook of patches around extracted interest points using clustering
2. Map the patch around each interest point to closest codebook entry
3. For each codebook entry, store all positions it was found, relative to object center [center is given]

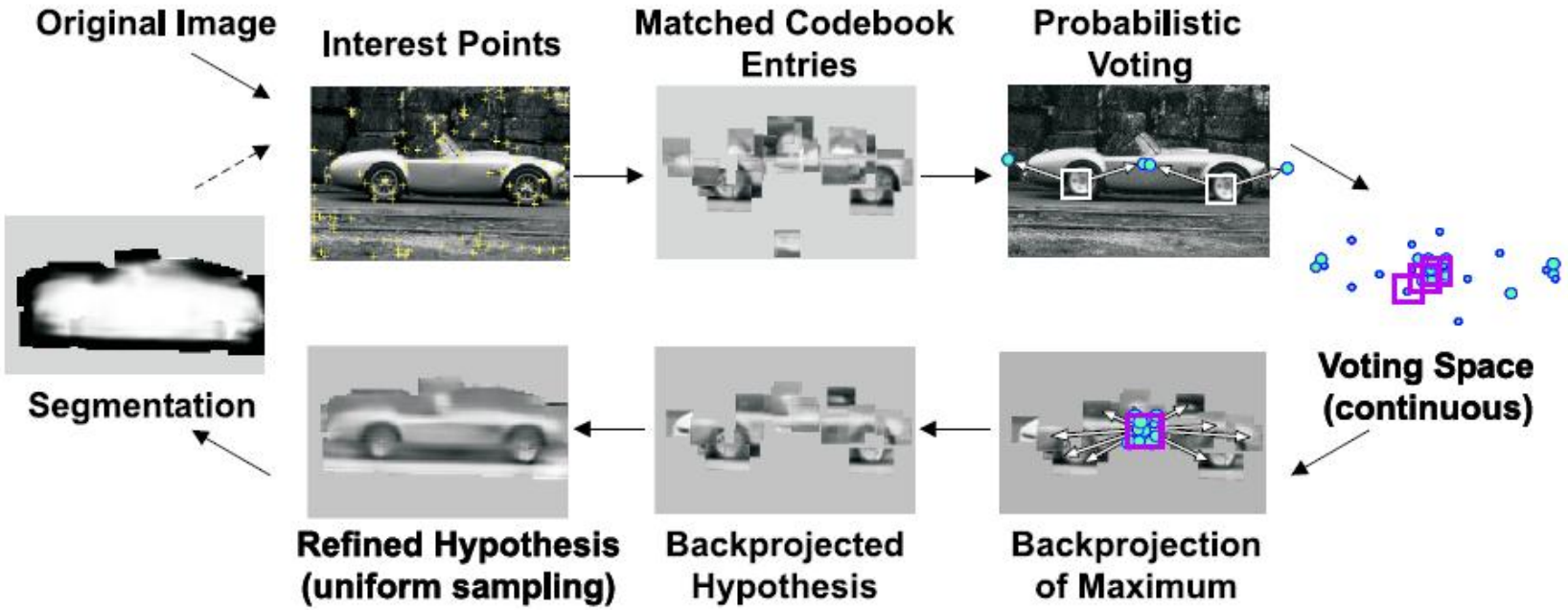




Credit slide: S. Lazebnik

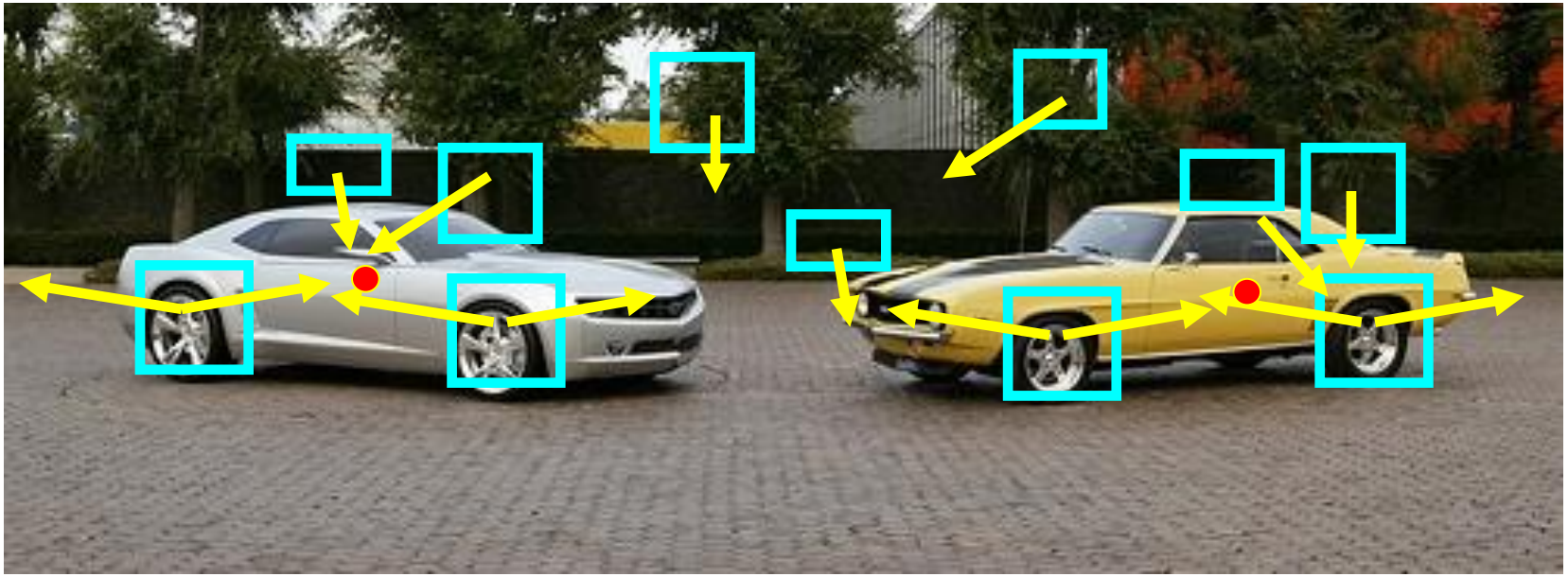
Implicit shape models: Testing

1. Given test image, extract patches, match to codebook entry
2. Cast votes for possible positions of object center
3. Search for maxima in voting space
4. Extract weighted segmentation mask based on stored masks for the codebook occurrences



Implicit shape models

- Visual codebook is used to index votes for object position [Cast votes for possible positions of object center]



test image

Search for maxima in voting space

Summary: Part based models

- Generative part-based models
  - Pro: very nice conceptually
  - Pro: semi-supervised!
  - Con: combinatorial hypothesis search problem

- Implicit shape models
  - Pro: can localize object, maintain translation and possibly scale invariance
  - Con: need supervised training data (known object positions and possibly segmentation masks)

Credit slide: S. Lazebnik

Outline

- Object Recognition

- Introduction
- Recognition of single 3D objects
  - Bag of world models
  - Part based models
  - Models for 3D objects categorization

3D Object Categorization



















- Weber et al. ‘00
- Schneiderman et al. ’01
- Capel et al ’02
- Johnson & Herbert ‘99

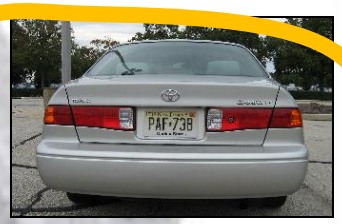
- Bronstein et al, ‘03
- Ruiz-Correa et al. ’03,
- Funkhouser et al ’03
- Bart et al ‘04

- Thomas et al. ‘06
- Kushal, et al., ’07
- Savarese et al, 07, 08

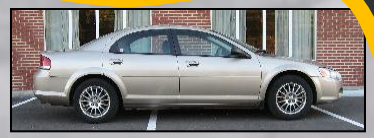
- Chiu et al.‘07
- Hoiem, et al., ’07
- Yan, et al. ’07

Single View Object Categorization
















- Leung et al '99
- Weber et al. '00
- Ullman et al. 02
- Fergus et al. '03
- Torralba et al. '03

- Felzenszwalb & Huttenlocher '03
- Fei-Fei et al. '04
- Leibe et al. '04

- Kumar & Hebert '04
- Sivic et al. '05
- Shotton et al '05
- Grauman et al. '05

- Sudderth et al '05
- Torralba et al. '05
- Lazebnik et al. '06
- Todorovic et al. '06
- Bosh et al '07
- Vedaldi & Soatto '08

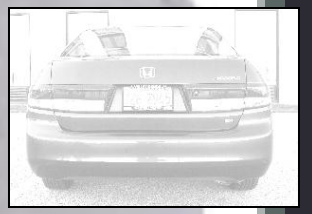
3D Object Categorization



















•Ballard, '81
•Grimson & L.-Perez, '87
•Lowe, '87

•Edelman et al. '91
•Ullman & Barsi, '91
• Rothwell '92
•Linderberg, '94
•Murase & Nayar '94

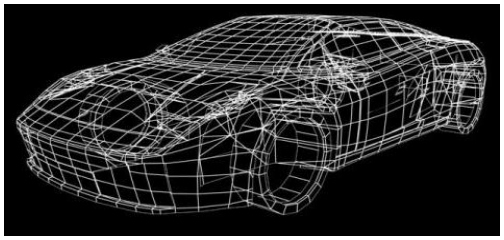
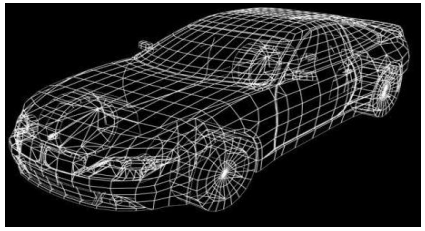
•Zhang et al '95
•Schmid & Mohr, '96
•Schiele & Crowley, '96
•Lowe, '99
•Jacob & Barsi, '99
•Rothganger et al., '04

•Ferrari et al, '05
•Brown & Lowe '05
•Snavely et al '06
•Yin & Collins, '07

3D Object Categorization

Additional challenges

- how to model 3D shape variability?

-How to model texture (appearance) variability?


3D Object Categorization

Mixture of 2D single view models

- Weber et al.'00
- Schneiderman et al.'01
- Bart et al.'04

Full 3D models

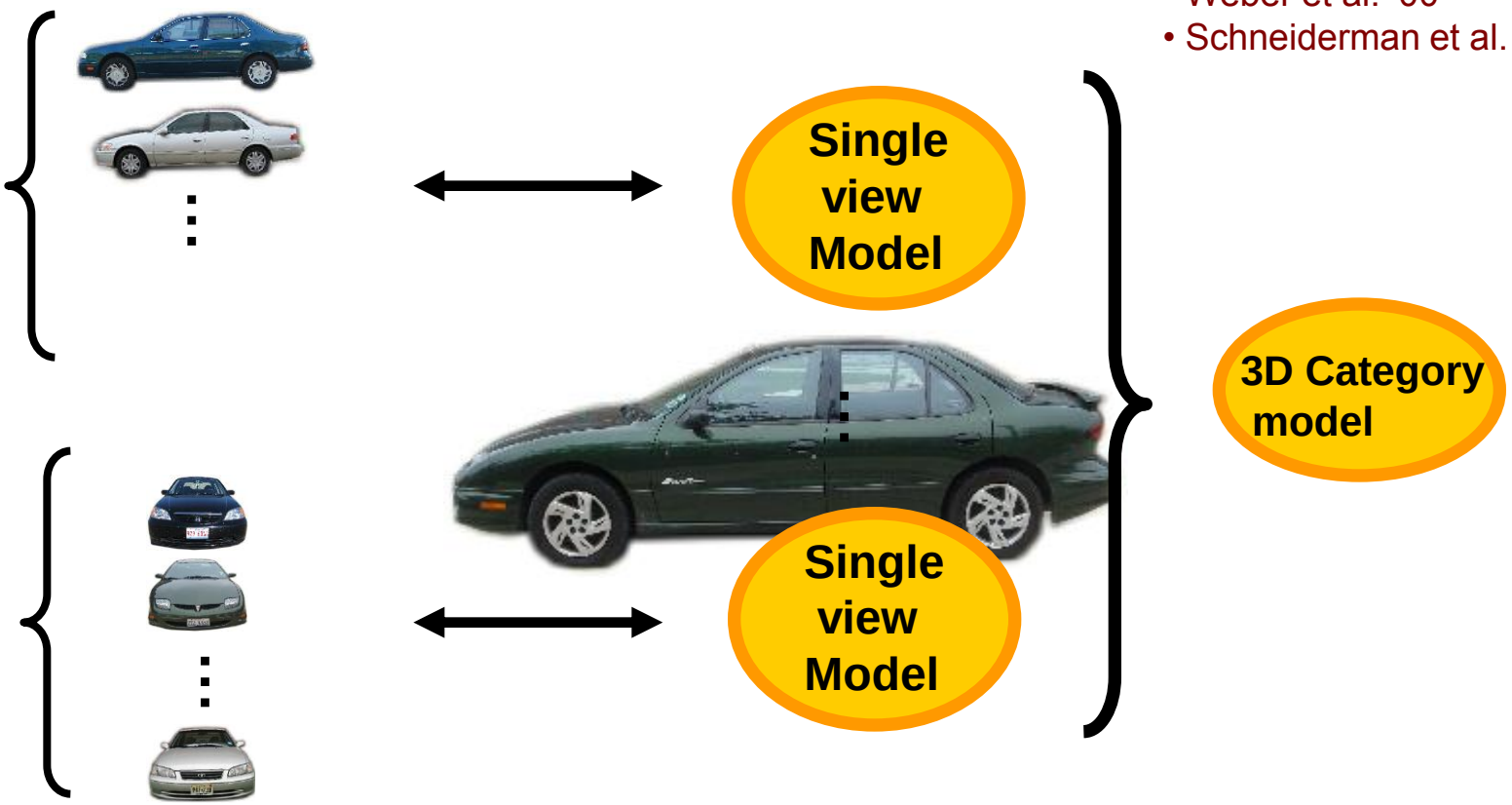
- Bronstein et al,'03
- Ruiz-Correa et al.'03,
- Funkhouser et al '03
- Capel et al '02
- Johnson & Herbert '99

Multi-view models

- Chiu et al.'07
- Hoiem, et al.,'07
- Yan, et al.'07

- Thomas et al.'06
- Kushal, et al.,'07
- Savarese et al, 07, 08

Mixture of single-view 2D models

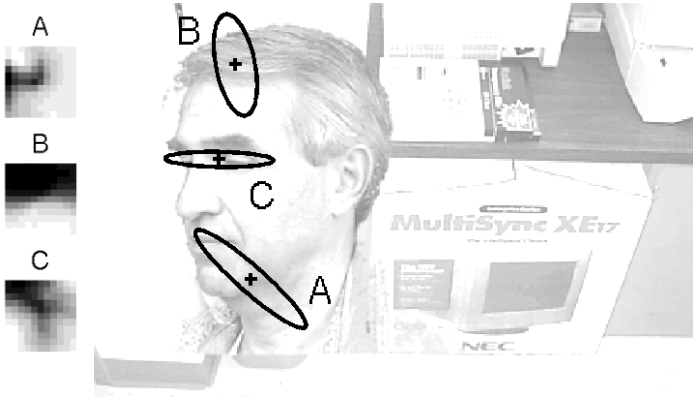


- Weber et al. ‘00
- Schneiderman et al. ’01

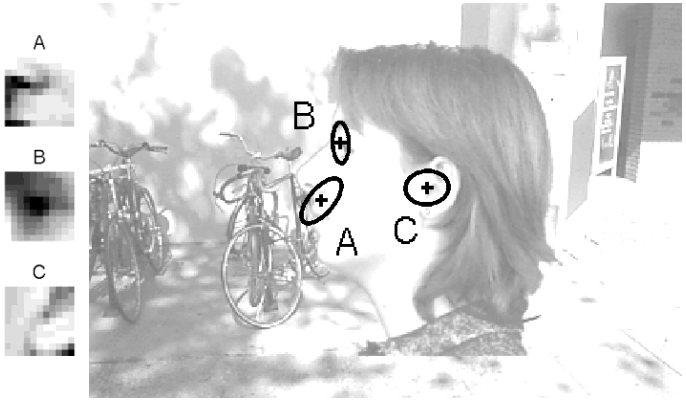
Multiple views

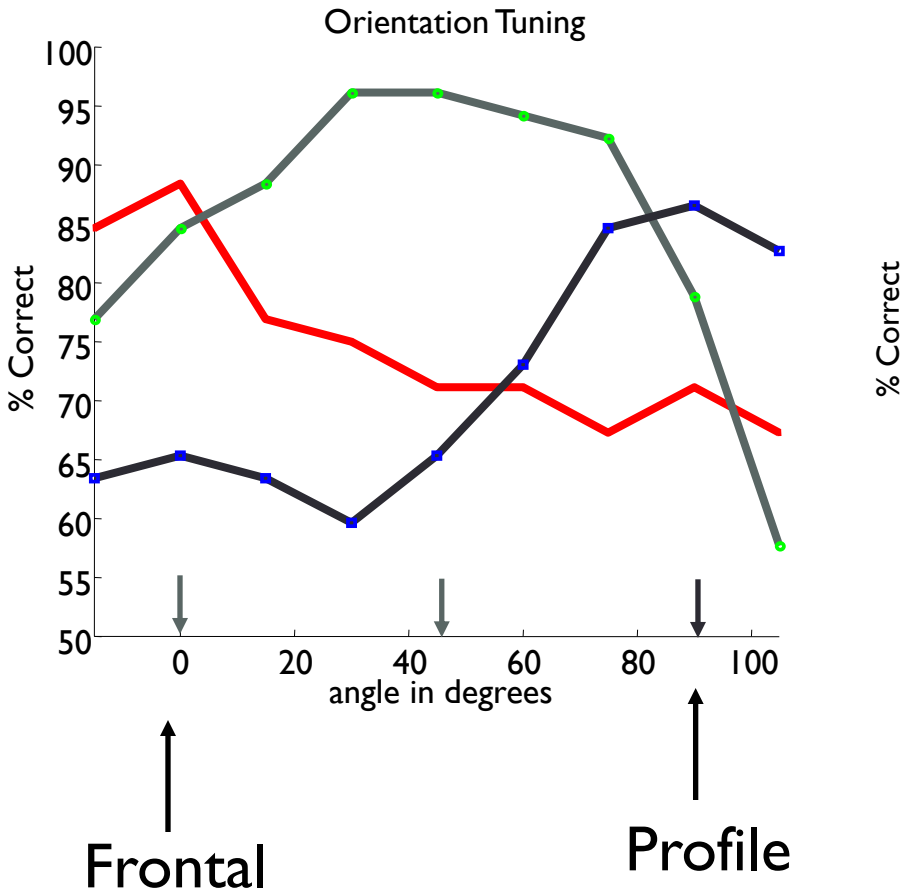
- Mixture of 2-D models
  - Weber, Welling and Perona CVPR '00

Component 1

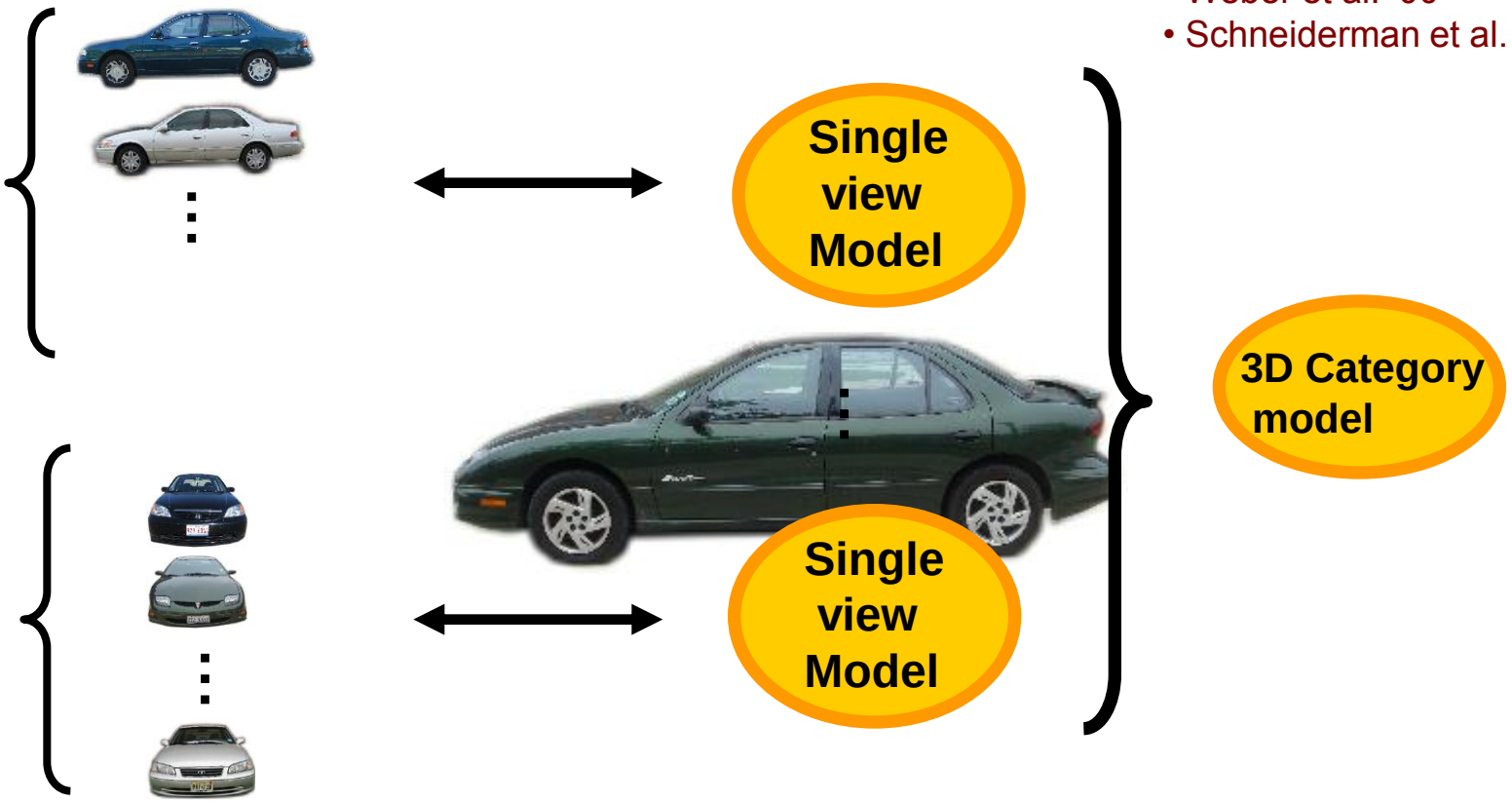


Component 2





Mixture of single-view 2D models



- Weber et al. ‘00
- Schneiderman et al. ’01

3D Object Categorization

Mixture of 2D single view models

- Weber et al.'00
- Schneiderman et al.'01
- Bart et al.'04

Full 3D models

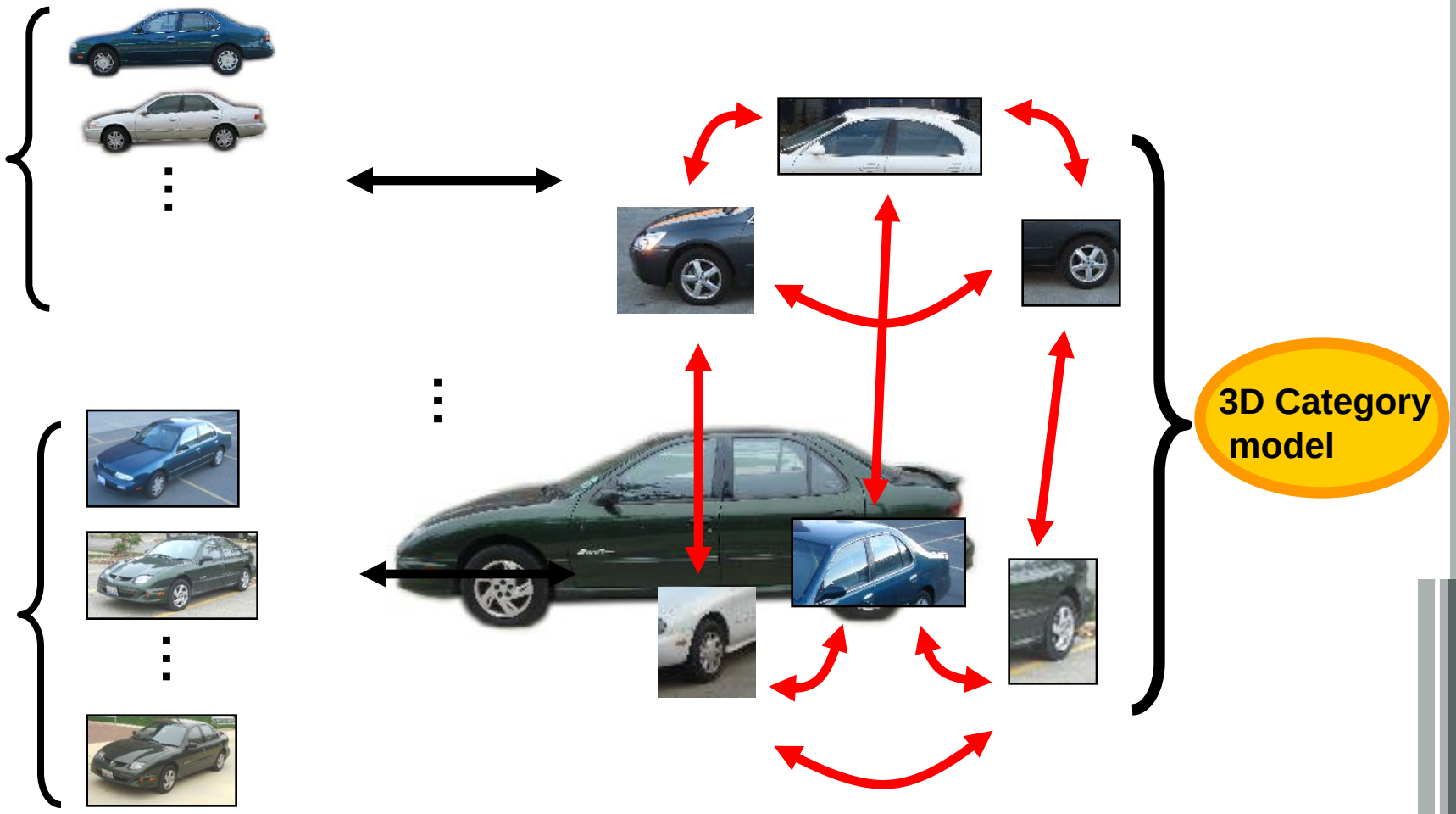
- Bronstein et al,'03
- Ruiz-Correa et al.'03,
- Funkhouser et al '03
- Capel et al '02
- Johnson & Herbert '99
- Chiu et al.'07
- Hoiem, et al.,'07
- Yan, et al.'07

Multi-view models

- Thomas et al.'06
- Kushal, et al.,'07
- Savarese et al, 07, 08

Multi-view models

- Thomas et al. '06
- Kushal, et al., '07
- Savarese et al, 07, 08

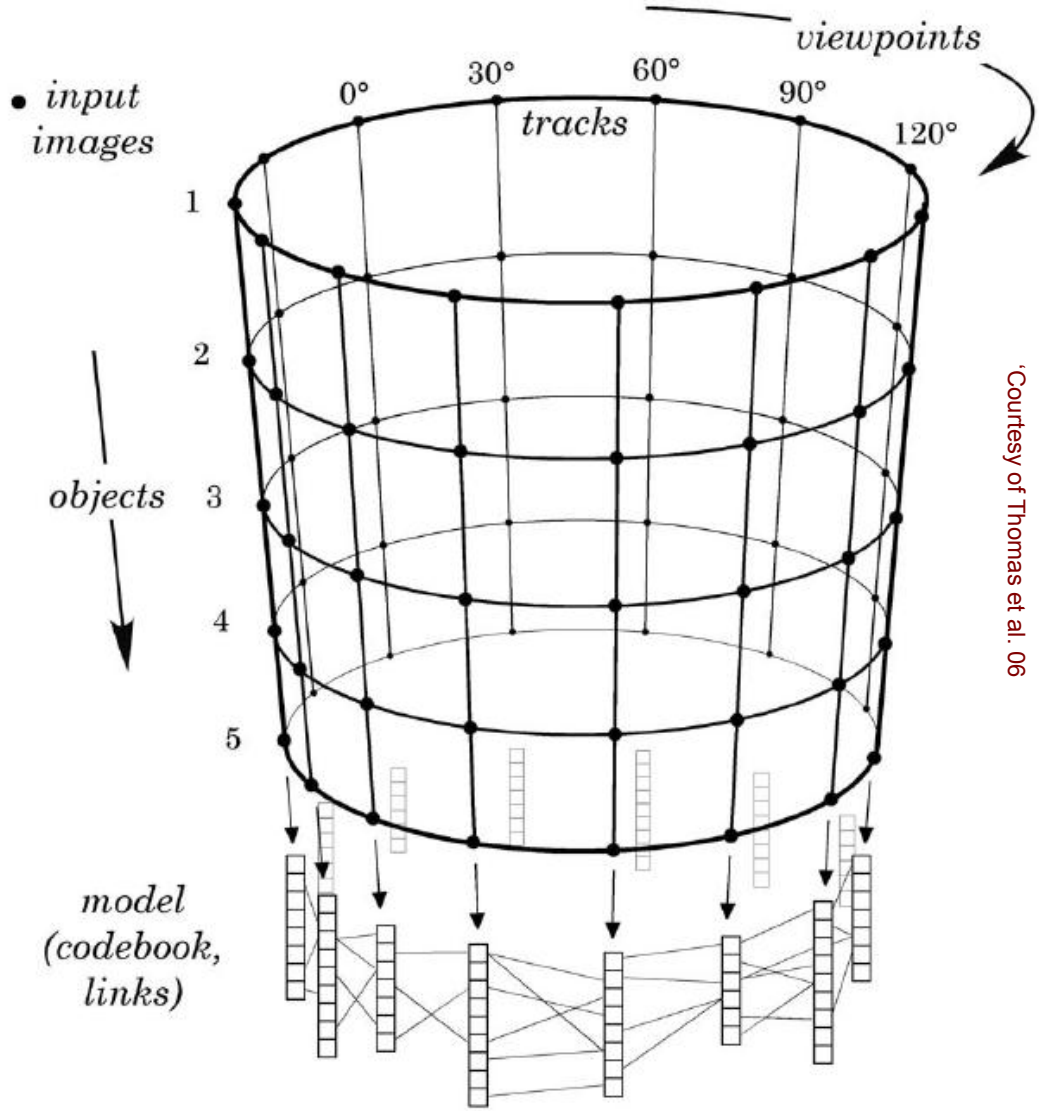


Sparse set of interest points or parts of the objects are linked across views.

Combining multi-views and ISM models

[Thomas et al. '06]

Representation

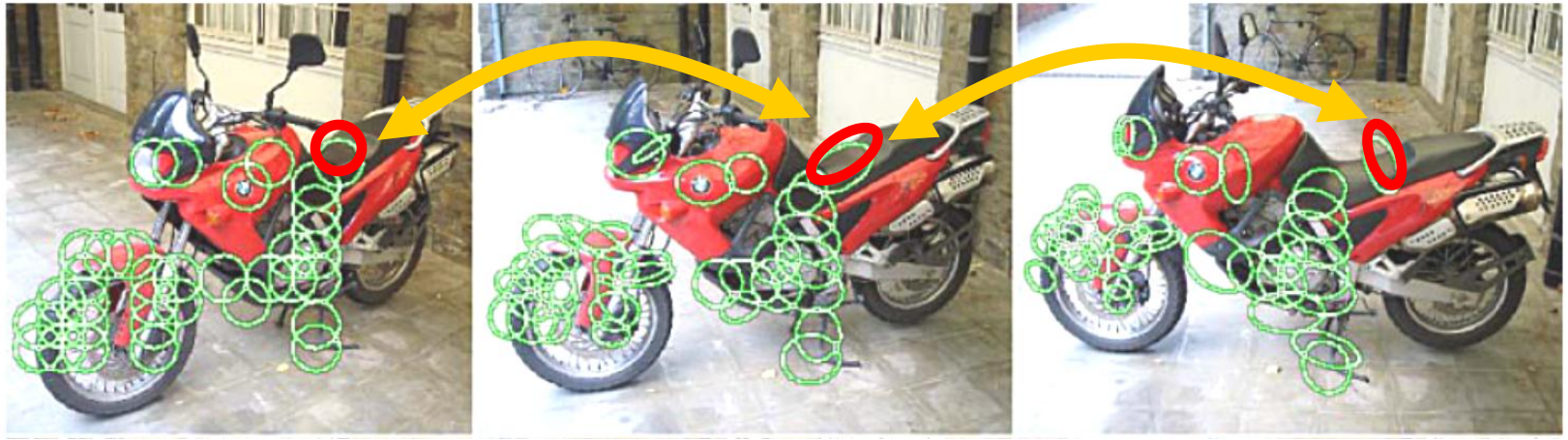


'Courtesy of Thomas et al. 06

Combining multi-views and ISM models

Representation

[Thomas et al. '06]
[Ferrari et al. '04, '06]
[Leibe et al. '04]



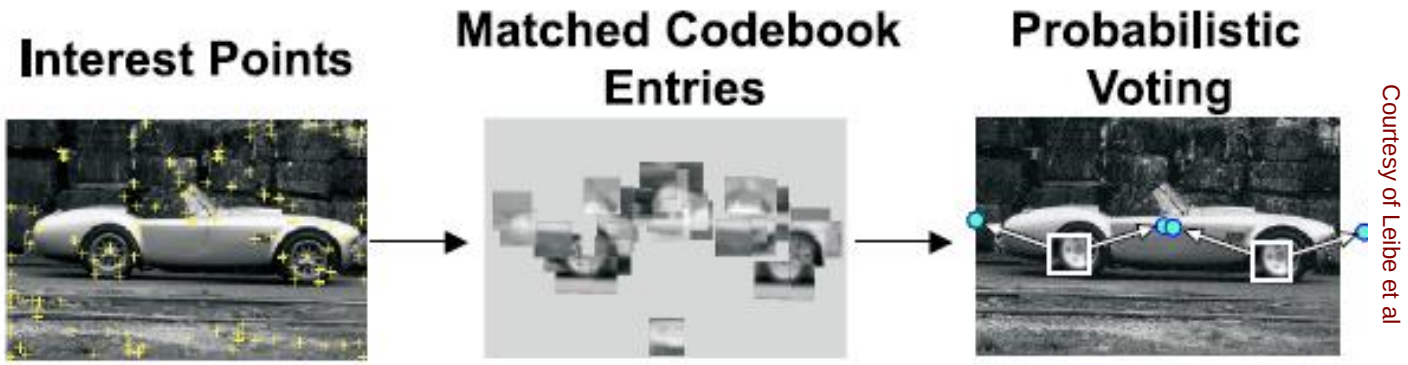
Courtesy of Thomas et al. 06

Set of *region-tracks* connecting model views
Each track is composed of image regions of a single physical surface patch along the model views in which it is visible.

[Ferrari et al. '04, '06]

Combining multi-views and ISM models

[Thomas et al. '06]



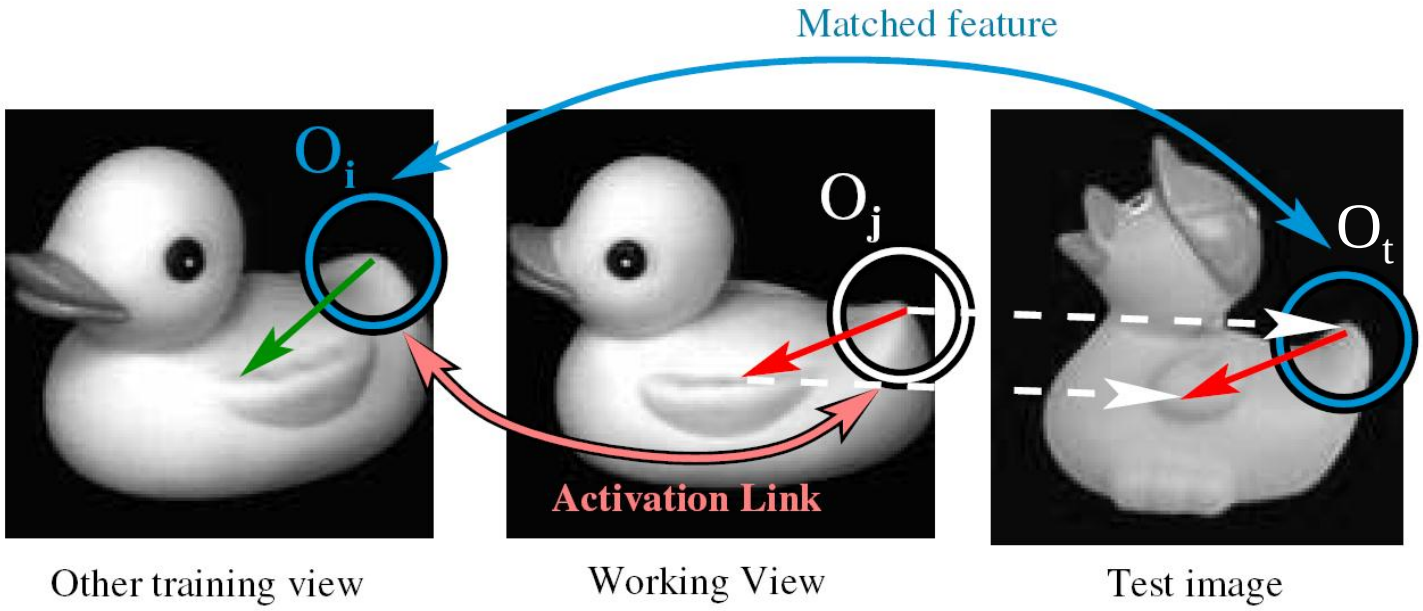
1. Features are extracted from the image
   … and matched to all the codebooks of the different ISMs.
2. Votes are cast in the Hough spaces of each ISM separately
3. Initial hypotheses are detected as local density maxima in these spaces.

bank of independent single-view detectors

Combining multi-views and ISM models

Recognition

[Thomas et al. '06]



Courtesy of Thomas et al. 06

4. A*ugment* the Hough spaces of each selected working view by inserting additional votes from codebook matches in other views [*vote transferring*]

- Qj is not being matched to Ot in the test image.
- However Qj is linked to Qi to another view via activation link;
- Since Qi is matched to Ot, an additional vote is added.
- This is vote transferring

Combining multi-views and ISM models

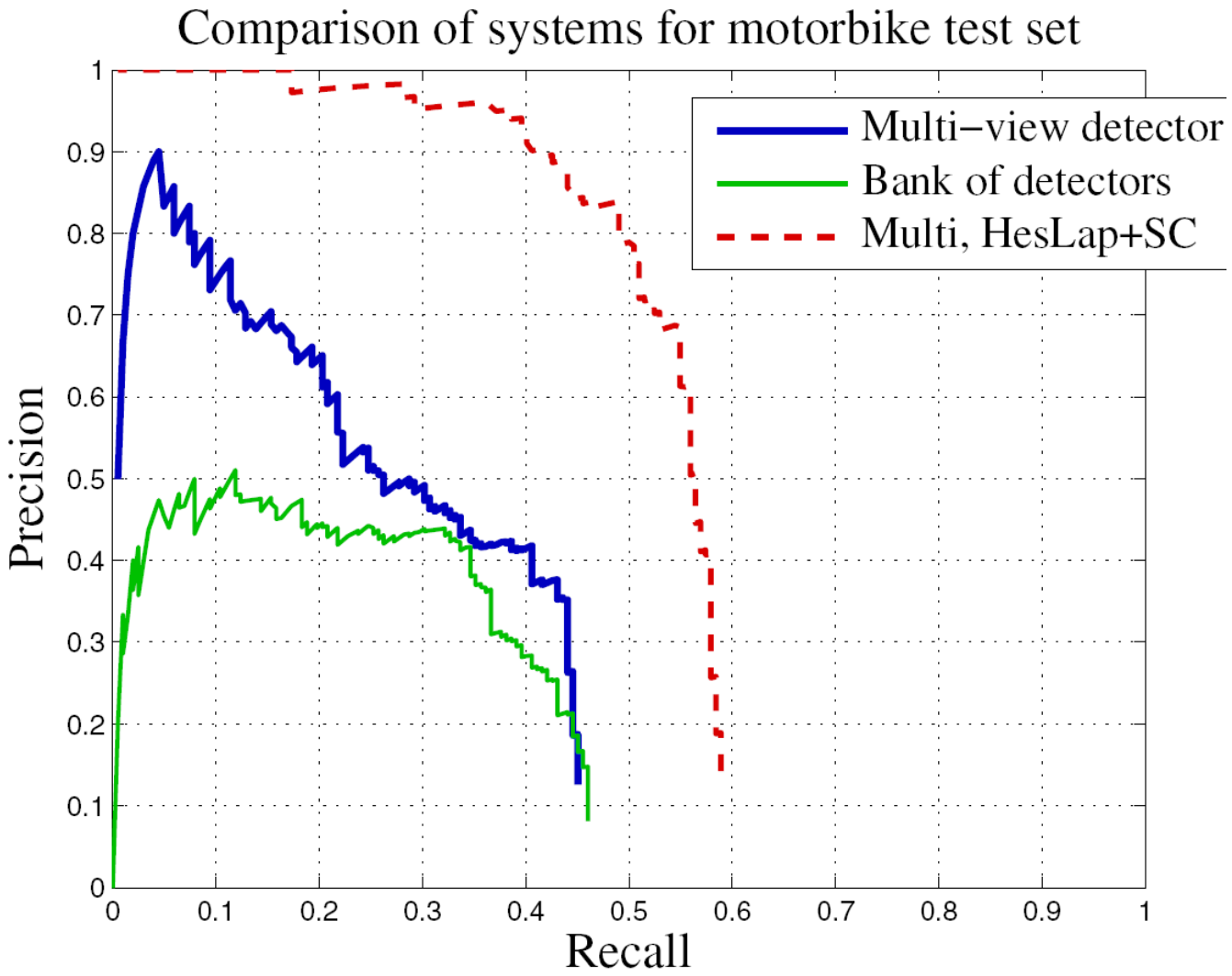
Results

[Thomas et al. ’06]

Combining multi-views and ISM models

Results

[Thomas et al. '06]



3D Object Categorization

Mixture of 2D single view models

- Weber et al.'00
- Schneiderman et al.'01
- Bart et al.'04

Full 3D models

- Bronstein et al,'03
- Ruiz-Correa et al.'03,
- Funkhouser et al '03
- Capel et al '02
- Johnson & Herbert '99
- Chiu et al.'07
- Hoiem, et al.,'07
- Yan, et al.'07

Multi-view models

- Thomas et al.'06
- Kushal, et al.,'07
- Savarese et al, 07, 08

Flexible models of object categories by PSMs

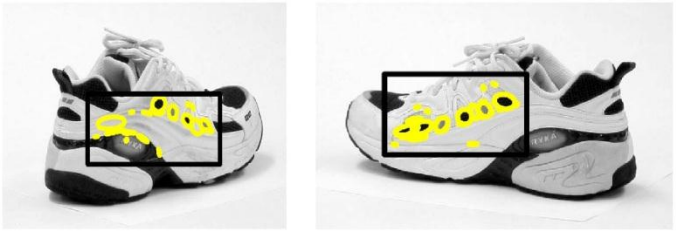
Representation

[Kushal et al. '06]

Courtesy of Kushal et al.

PSMs = dense, locally rigid assemblies of texture patches

• PSMs are learned by matching repeating patterns of features across training images of each object class



[Lazebnick et al '04]

3D Object Categorization

Mixture of 2D single view models

- Weber et al.'00
- Schneiderman et al.'01
- Bart et al.'04

Full 3D models

- Bronstein et al,'03
- Ruiz-Correa et al.'03,
- Funkhouser et al '03
- Capel et al '02
- Johnson & Herbert '99
- Chiu et al.'07
- Hoiem, et al.,'07
- Yan, et al.'07

Multi-view models

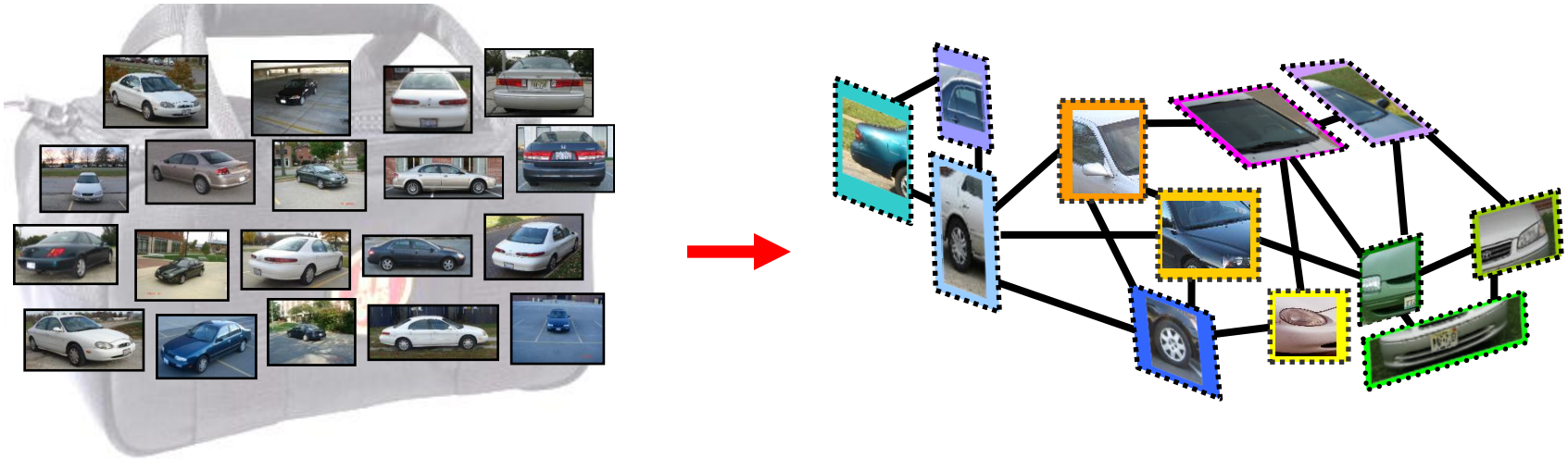
- Thomas et al.'06
- Kushal, et al.,'07
- Savarese et al, 07, 08

Linkage structure of canonical parts

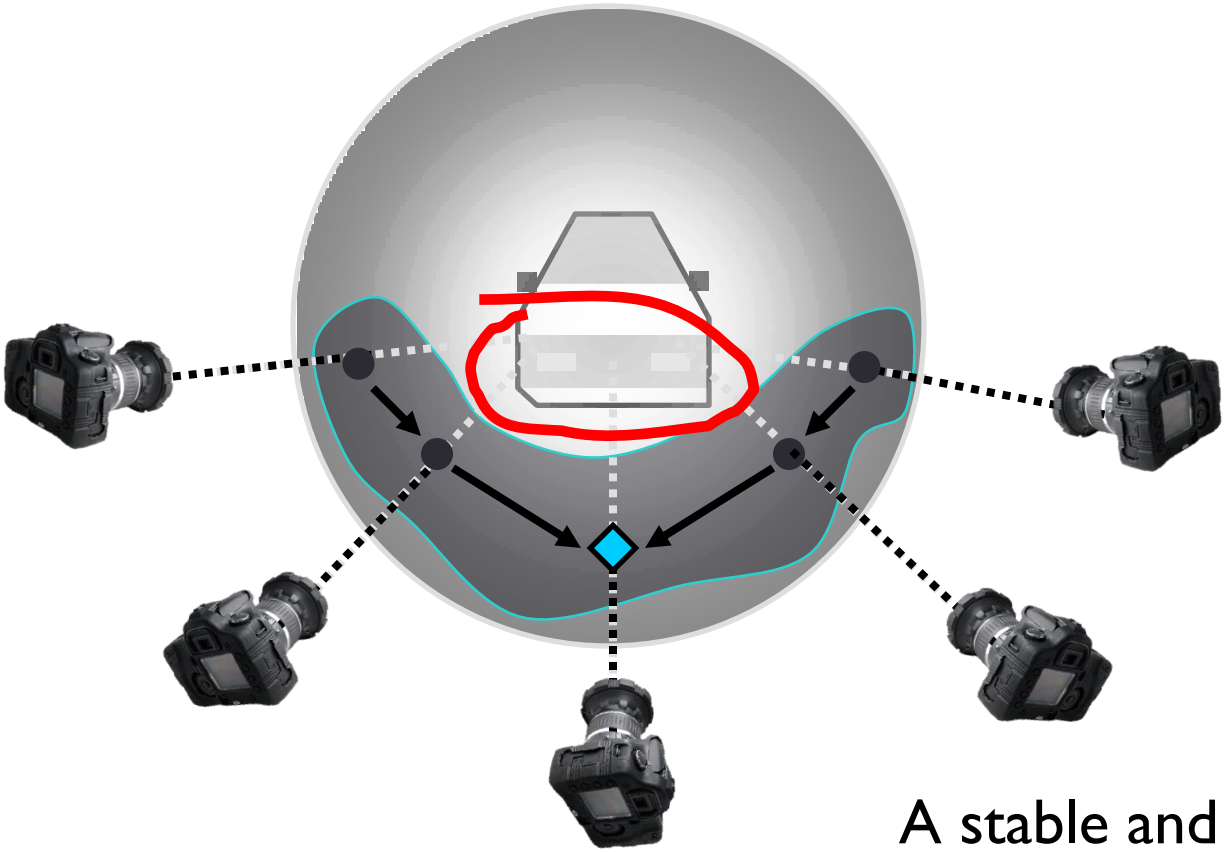
Representation

Savarese & Fei-Fei, ‘07, ‘08

- **Canonical parts**
- **Linkage structure via weak geometry**

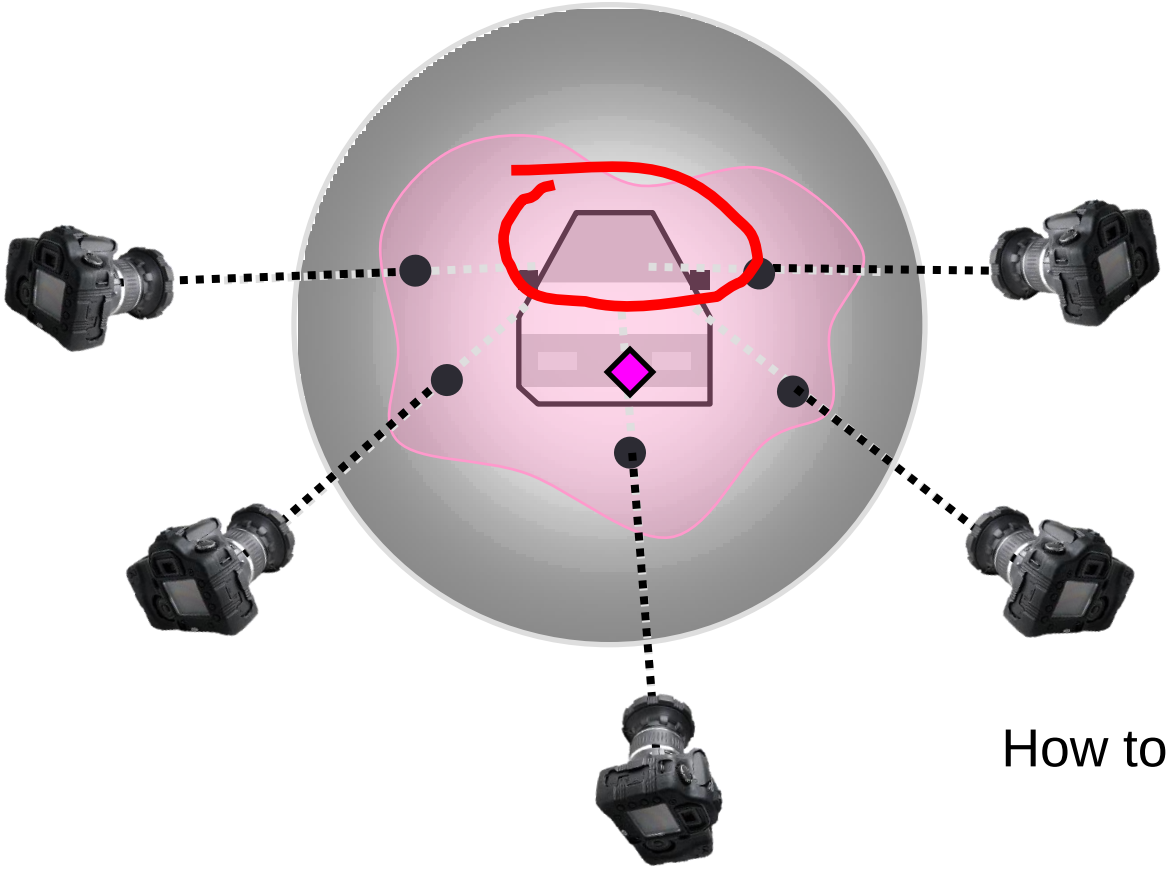


Canonical parts



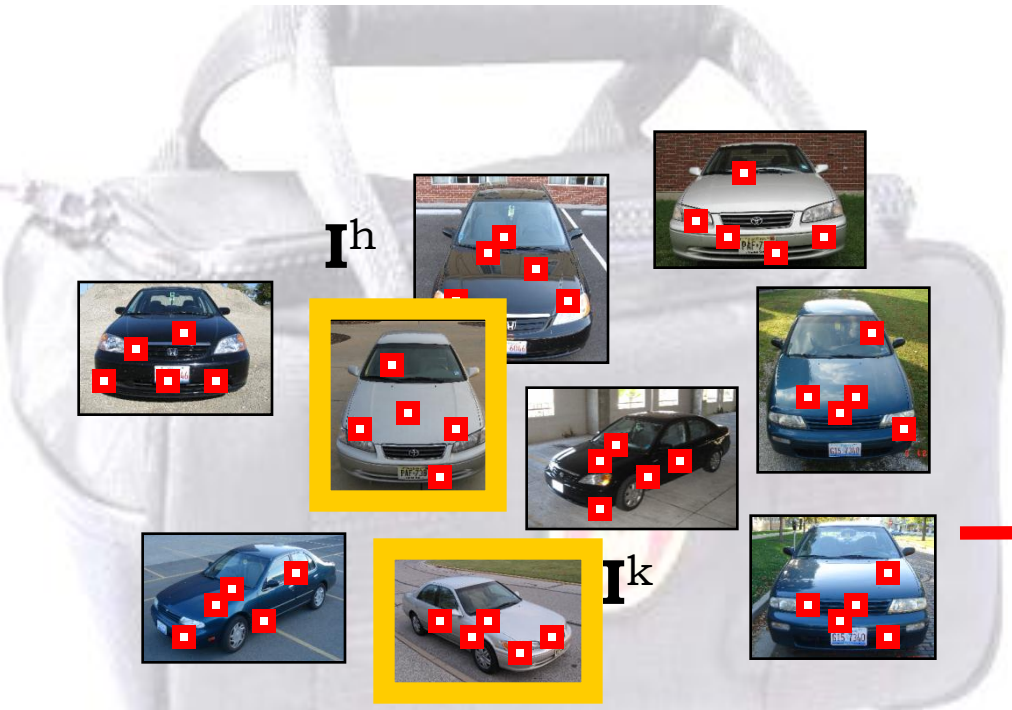
A stable and compact representation across multiple views!

Canonical parts

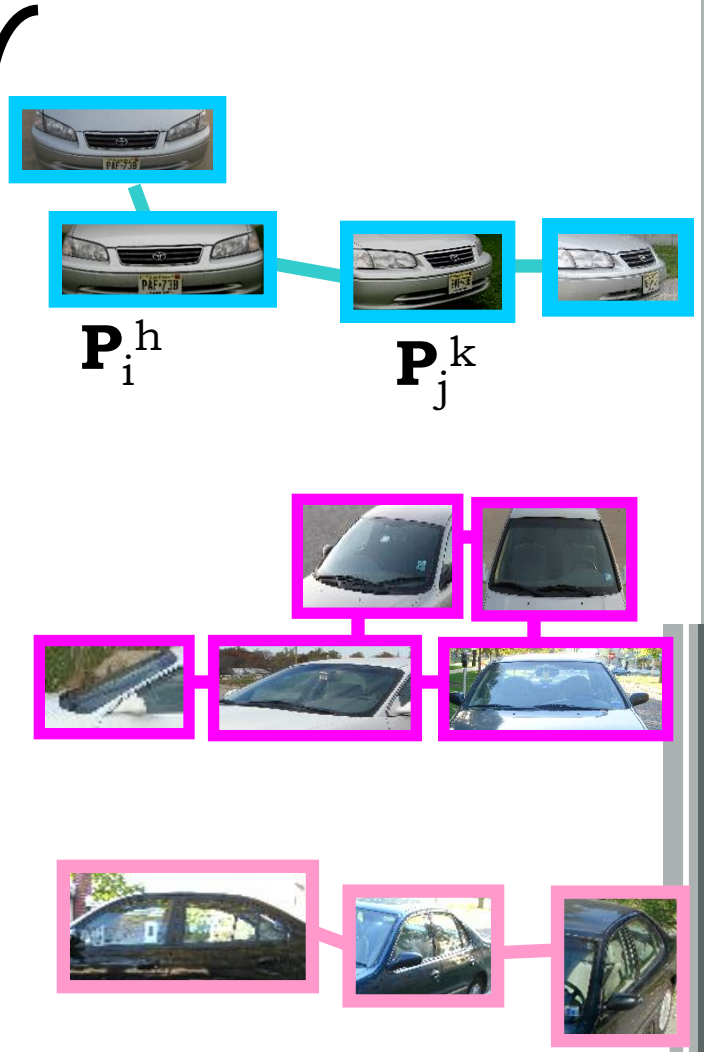


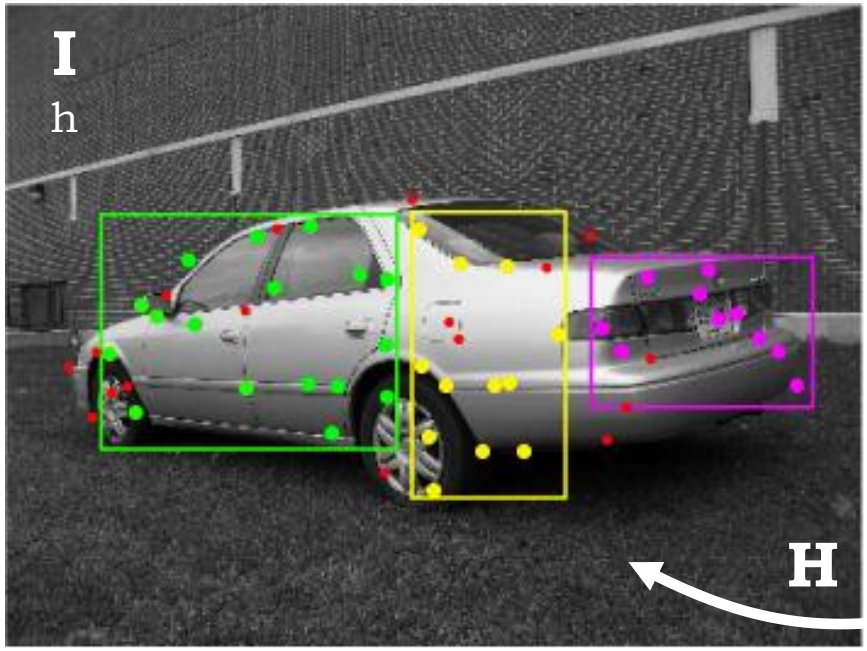
How to learn canonical parts?

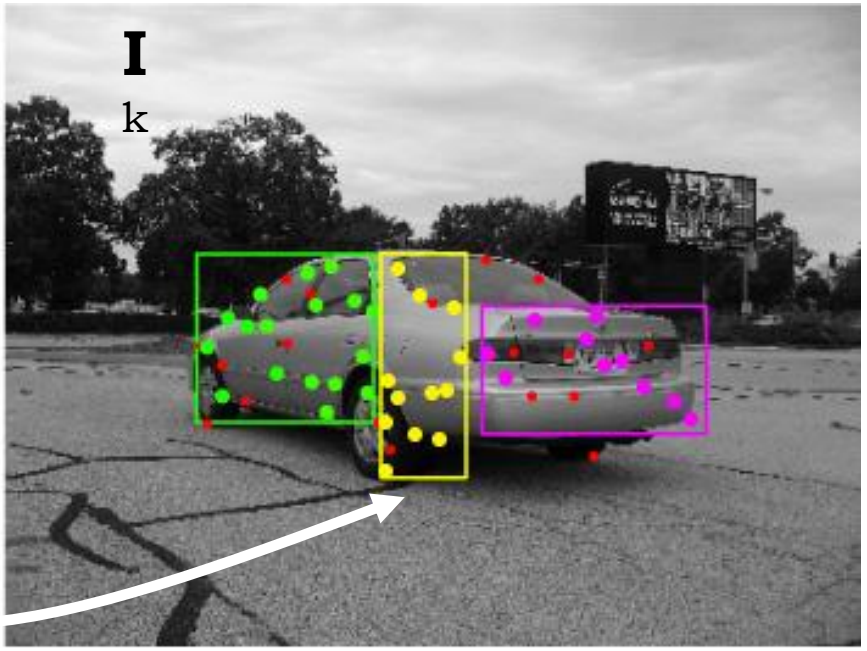
Canonical parts



Unlabeled mix of images





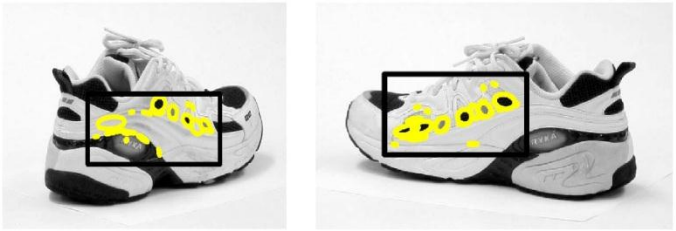


$$\pi : I^h \rightarrow \left\{ P_1^h, P_2^h, P_3^h, O^h \right\}$$

$$\tau : I^k \rightarrow \left\{ P_1^k, P_2^k, P_3^k, O^k \right\}$$

- Match candidates based on appearance
- Use RANSAC on H & F to filter best matches & partition matched features into matched parts
- output: matched parts related by H

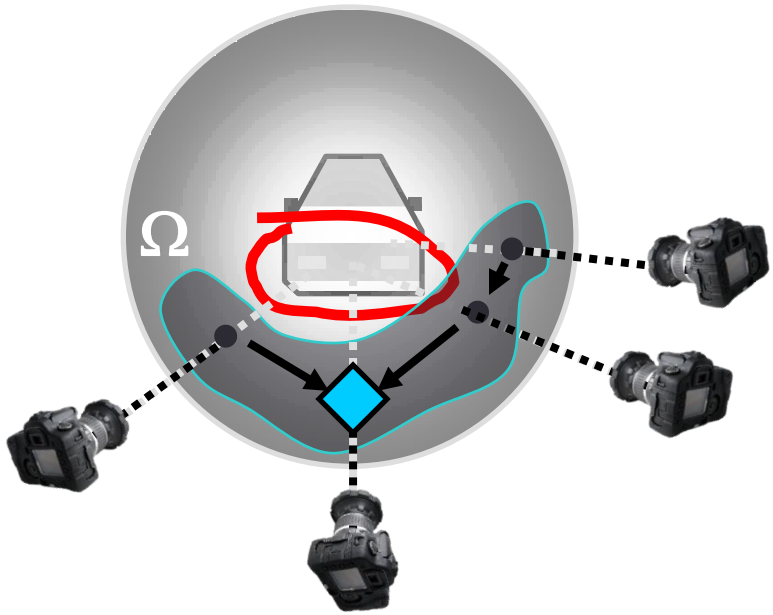
Alternative methods

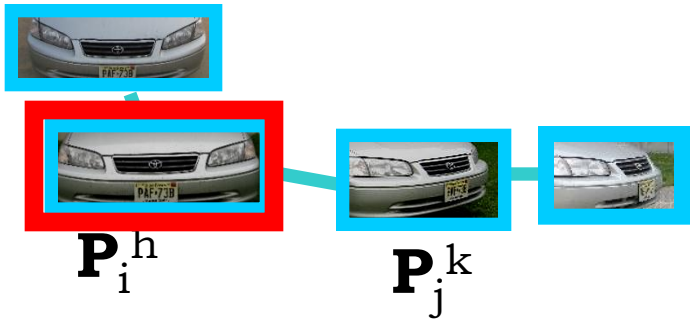


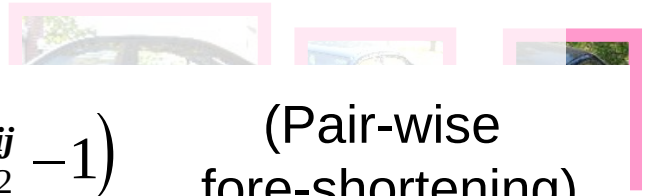
[Lazebnick et al '04]

[Ferrari et al '04]

Canonical parts



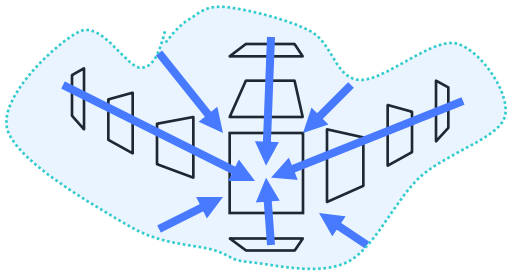




Property:

1. A canonical part is a stable point in the manifold if physical part is planar
2. The stable point can be reached by descending the gradient ∇

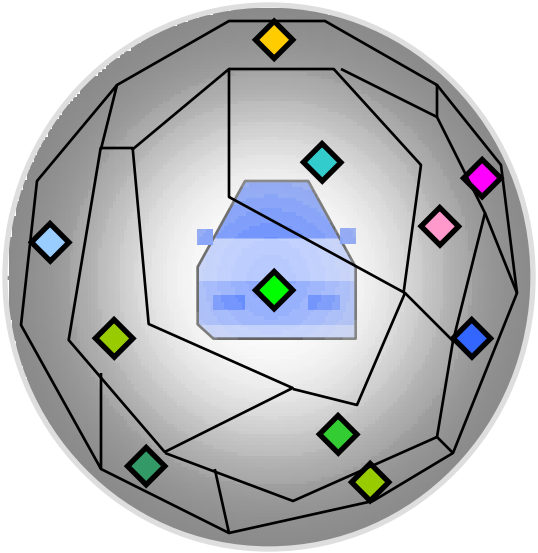
Ω

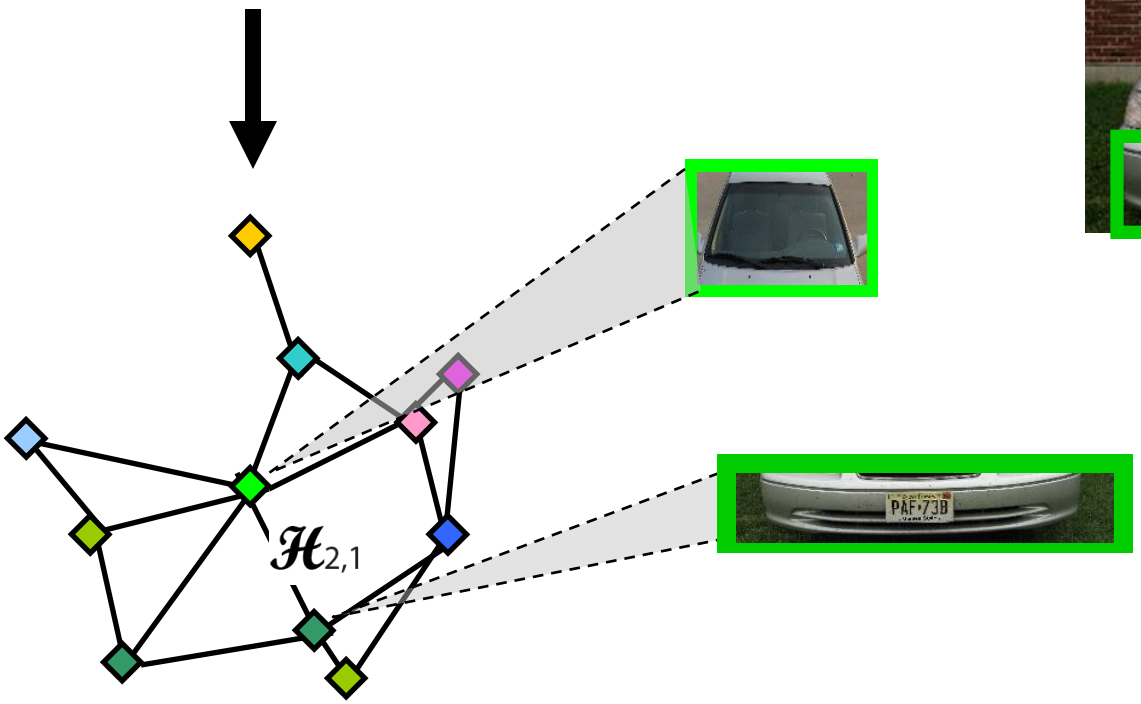


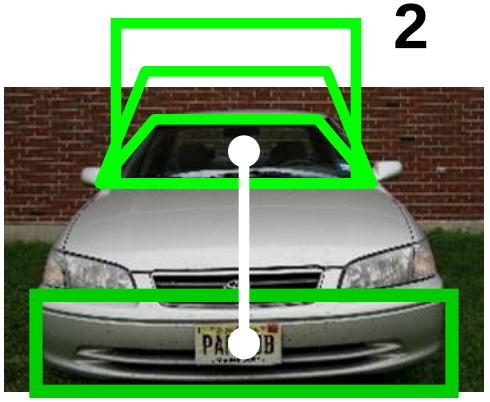
$$\|\nabla\|_{ij} \cong \left(\lambda_1^{ij}\lambda_2^{ij} - 1\right)$$

(Pair-wise fore-shortening)

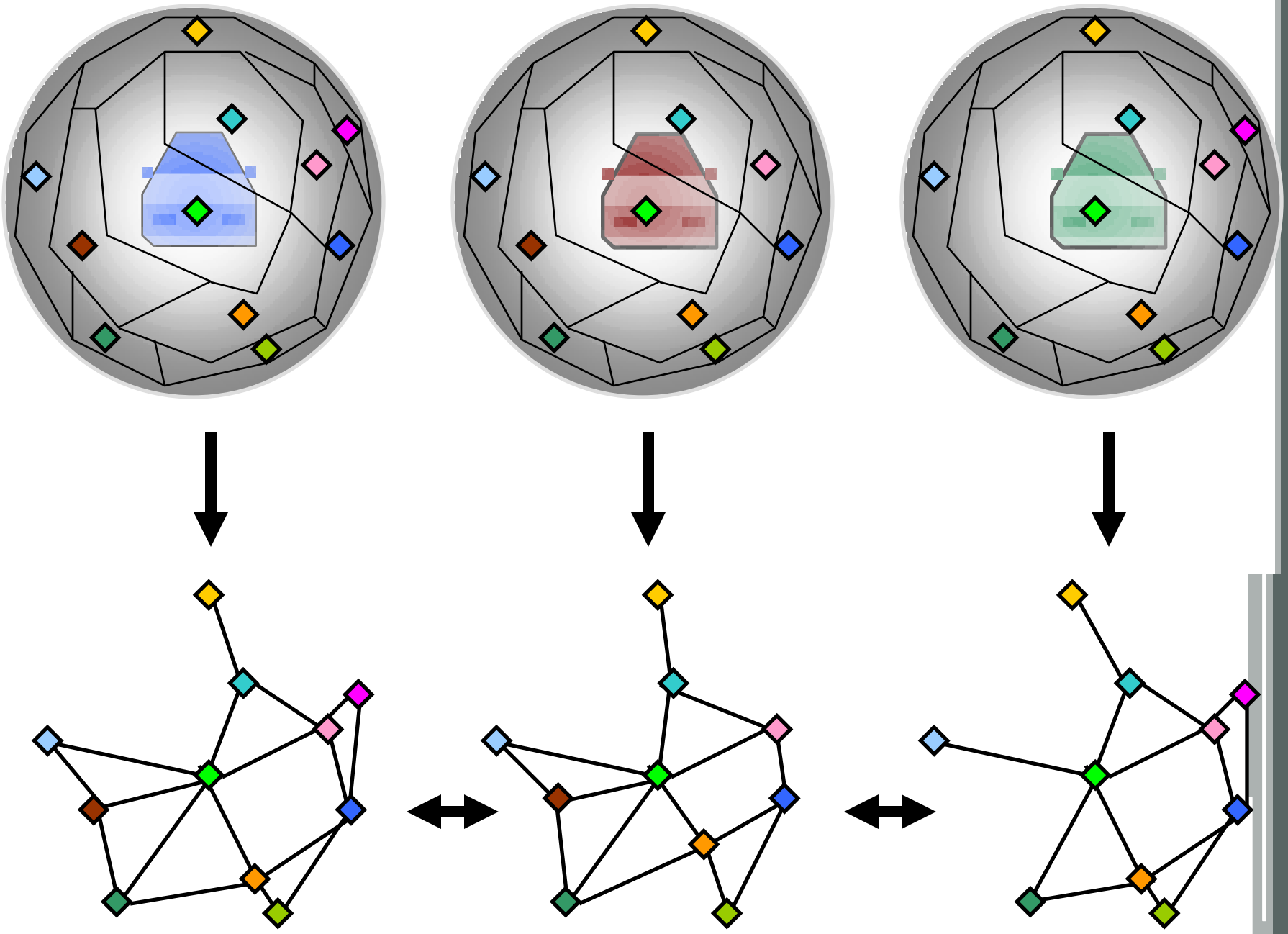
Linkage structure

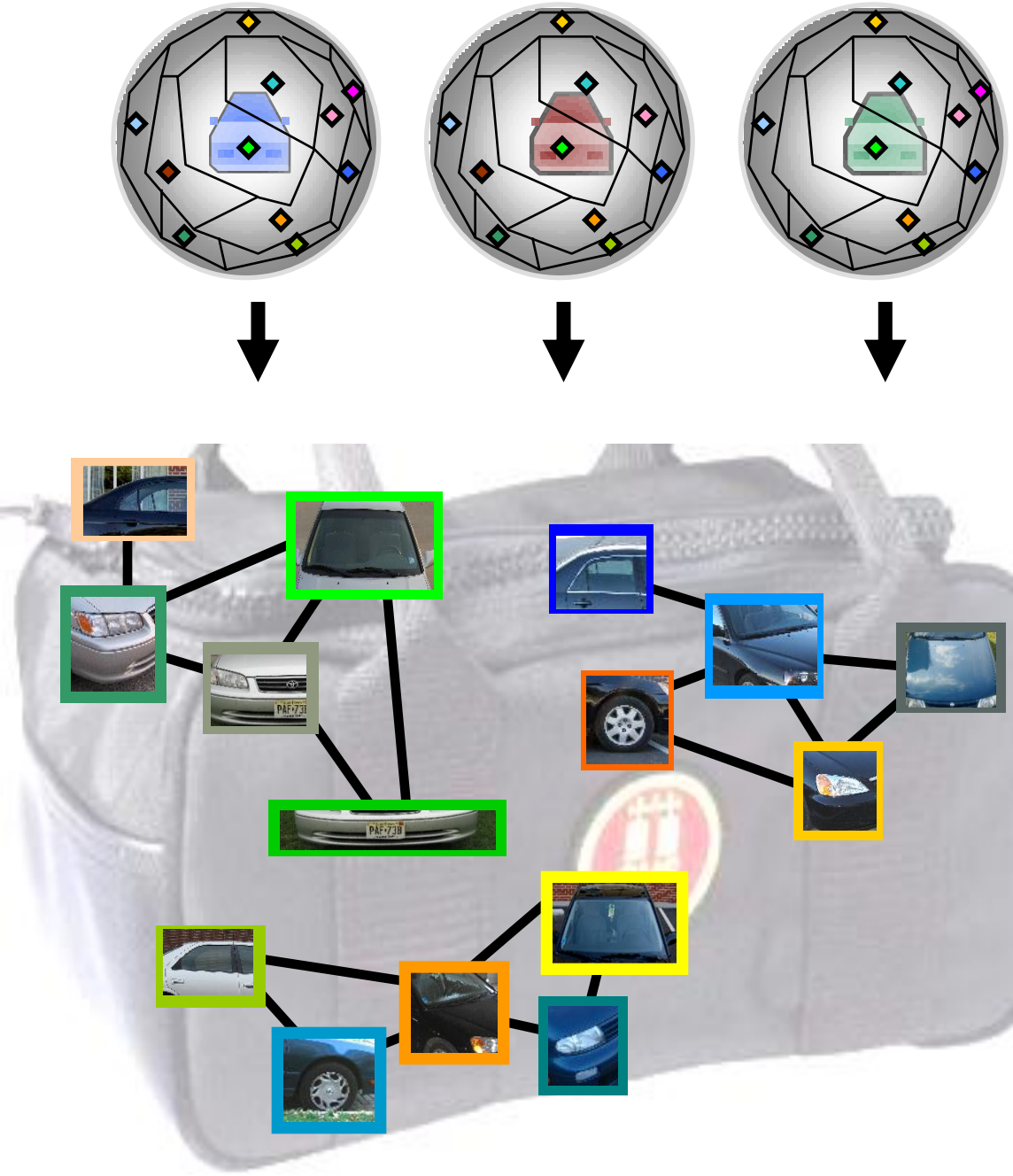


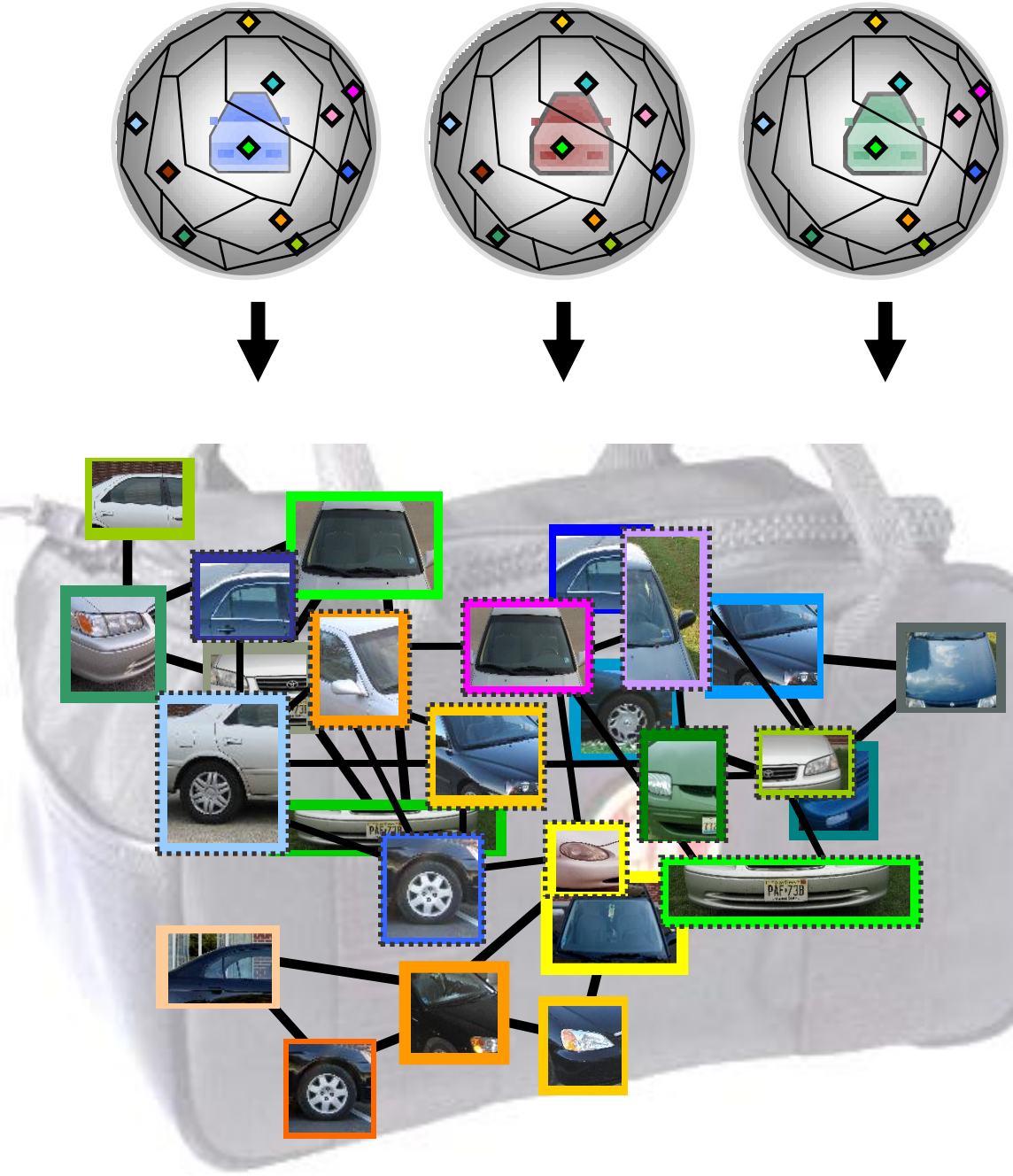




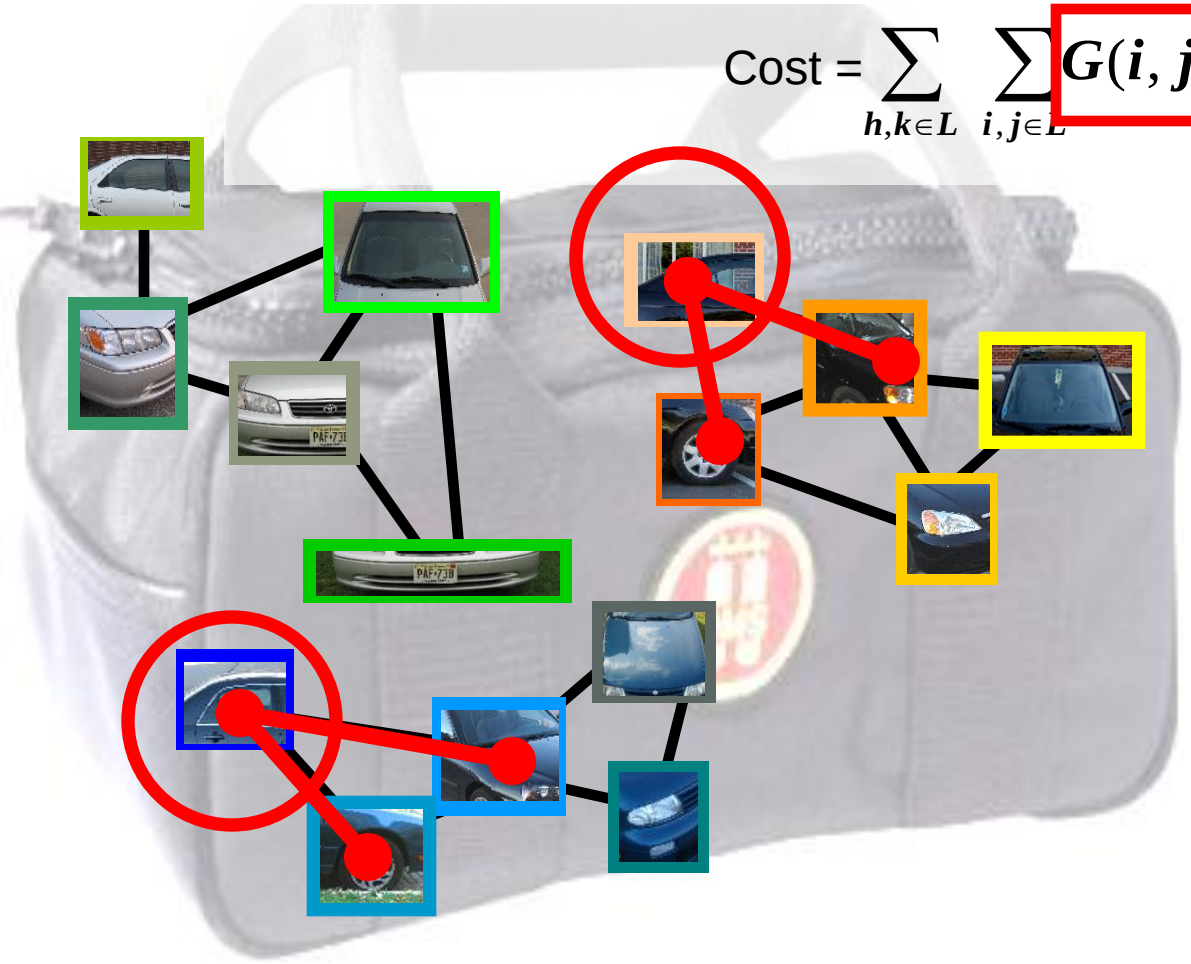
$$\mathcal{H}_{2,1} = \begin{pmatrix} \mathbf{H}_{2,1} & \mathbf{t}_{2,1} \\ \mathbf{0} & 1 \end{pmatrix}$$







Category Model



$$\text{Cost} = \sum_{h,k \in L} \sum_{i,j \in L} G(i,j,h,k)\,\delta_{ij}\delta_{hk} + \sum_{i,j \in L} A(i,j)\,\delta_{ij}$$

$$\sum_{j \in L} \delta_{ij} = 1 \quad \forall\, i$$

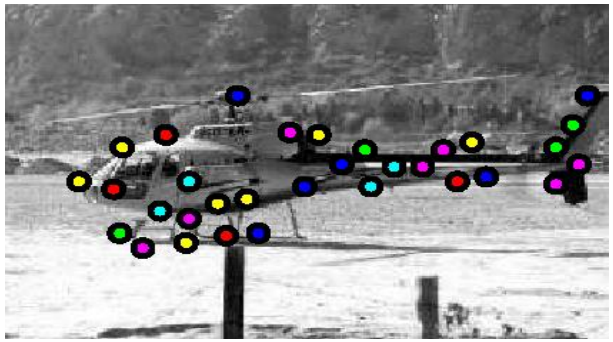
$$\delta_i = \{0,1\}$$

IQP problem
NP-complete $\rightarrow$ $O(N^2)$

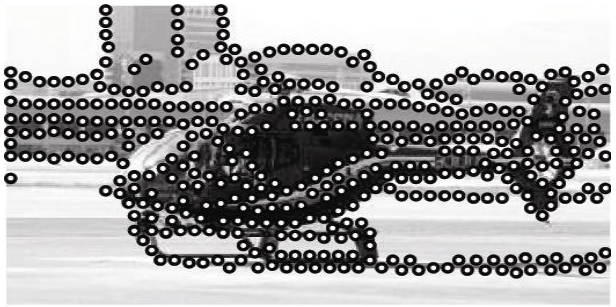
Maciel & Costeira '03
Berg et al '05
Leordeanu & Hebert '05

Deformable Template Matching

Berg, Berg and Malik CVPR 2005



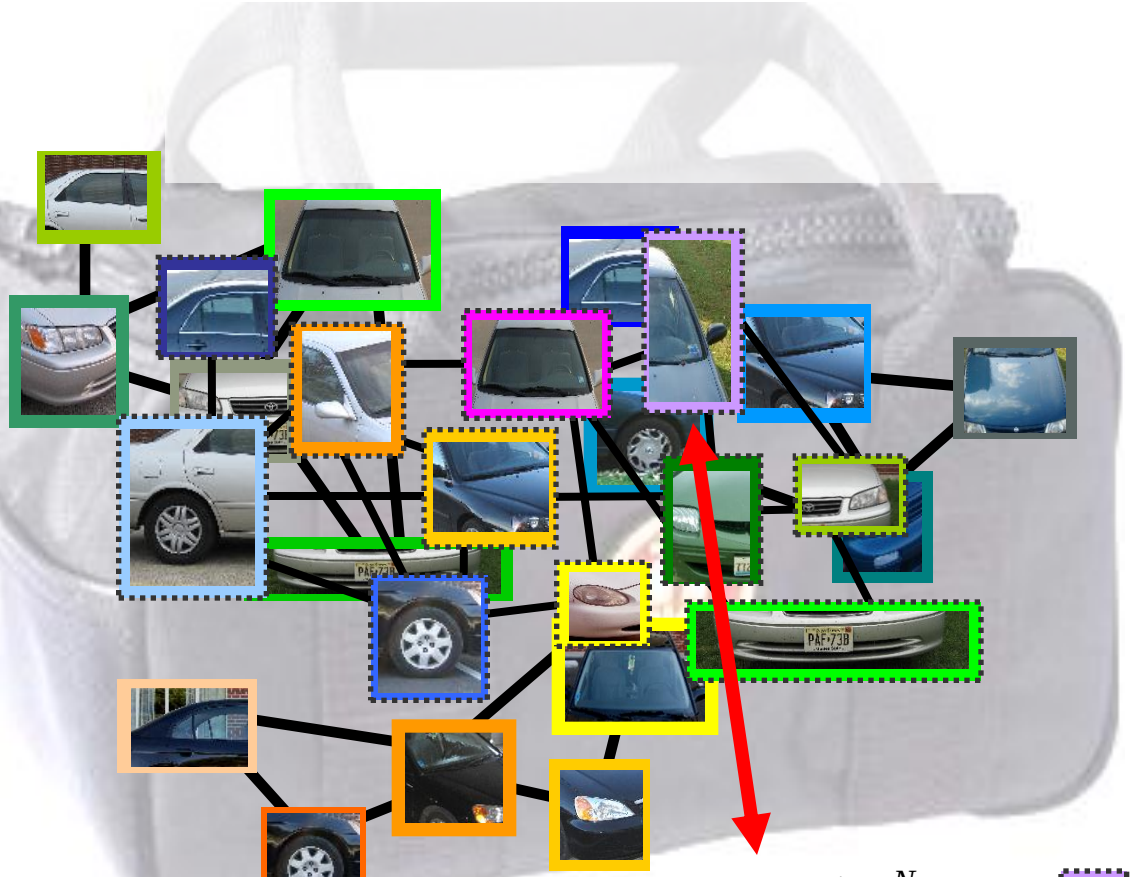
Template



Query

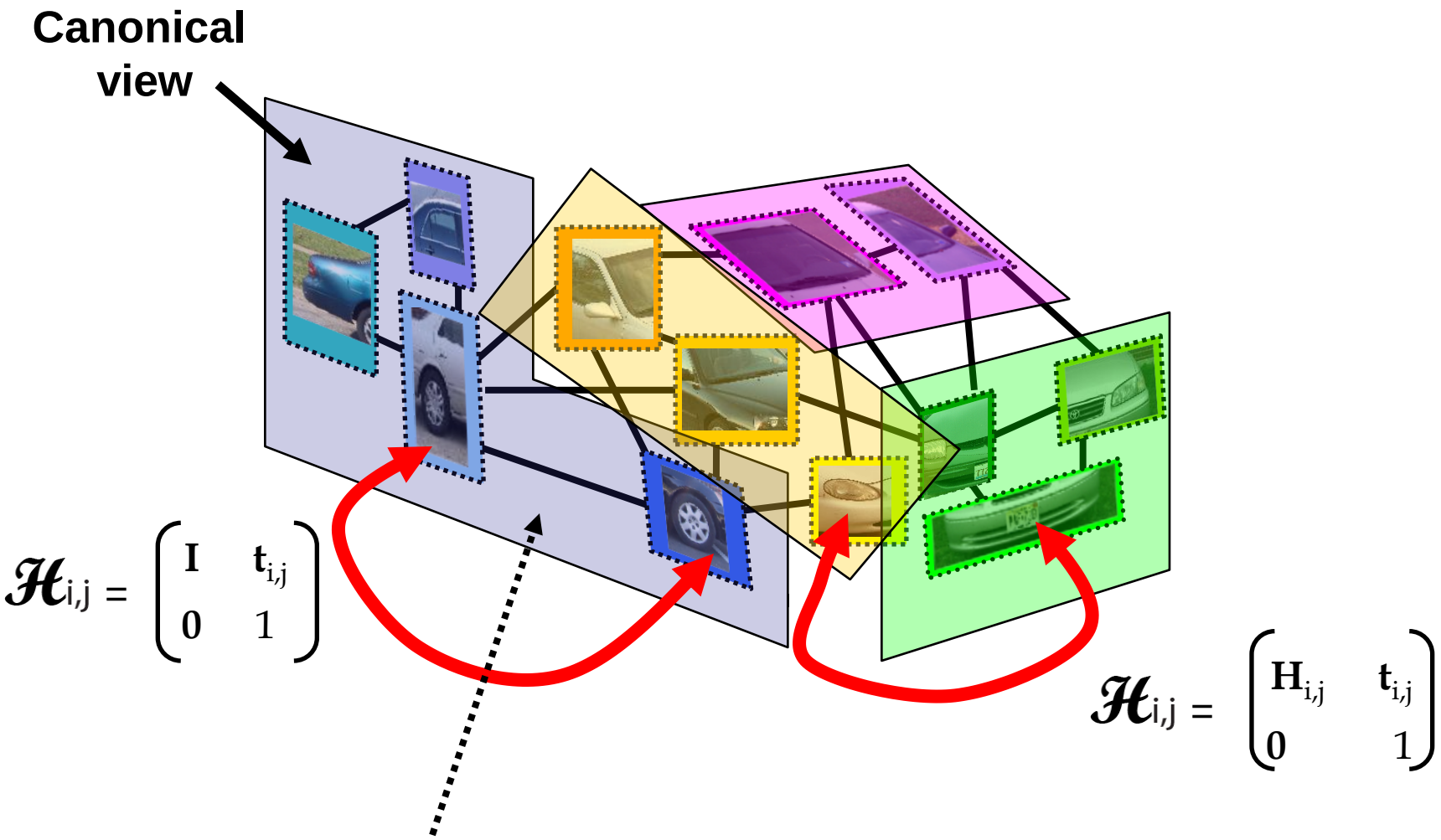
- Formulate problem as Integer Quadratic Programming
- $O(N^P)$ in general
- Use approximations that allow P=50 and N=2550 in <2 secs

Category Model

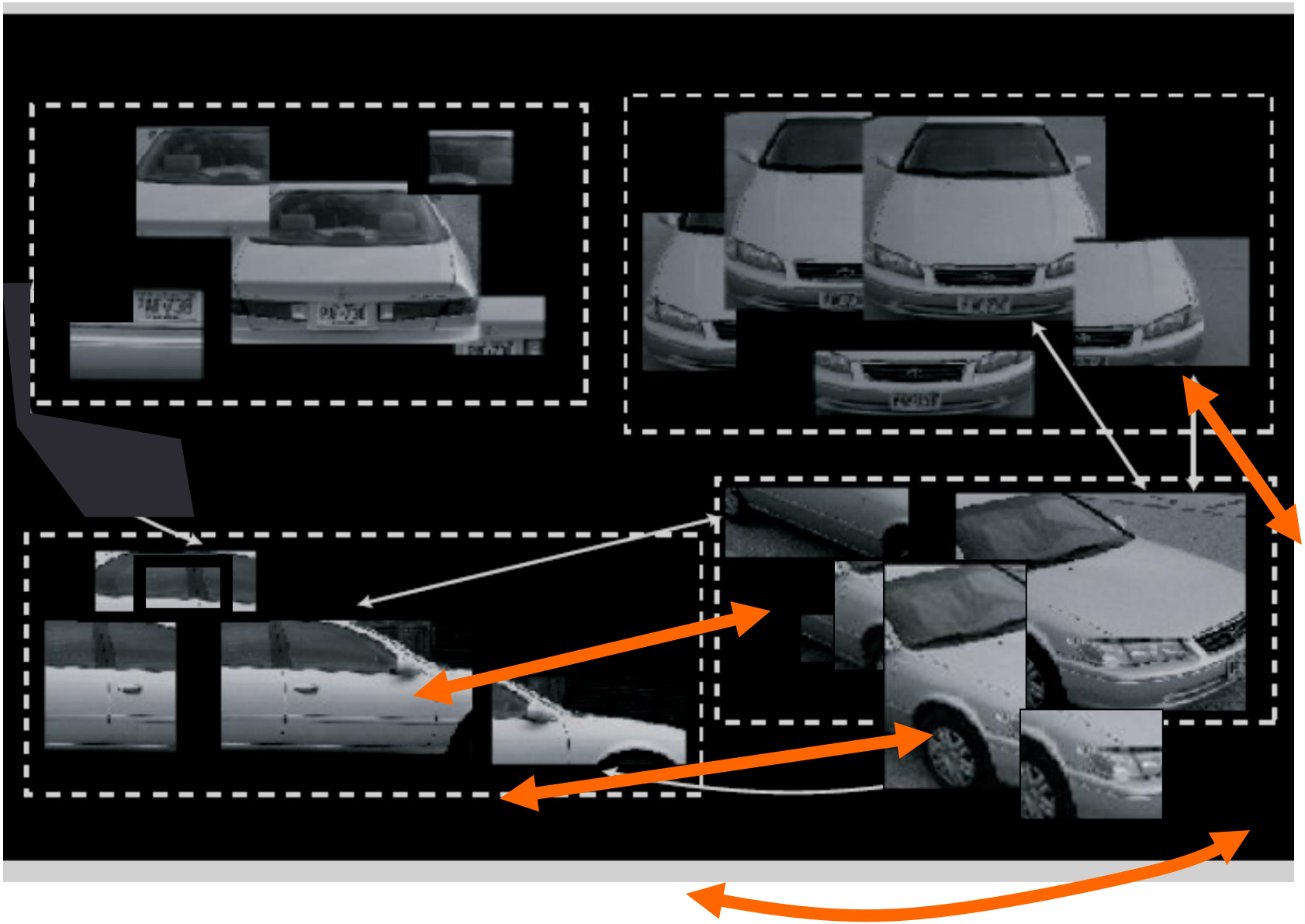


$$P_j = \frac{1}{N}\sum_{k}^{N} P_j^k$$

Category Model



2D constellation model!



Recognition

Query image

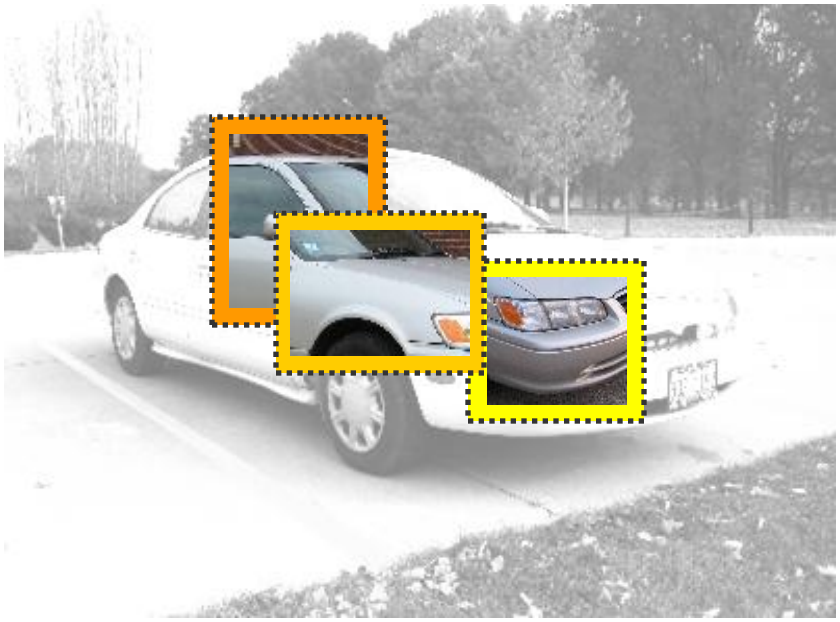
model

Algorithm

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

Recognition

Query image



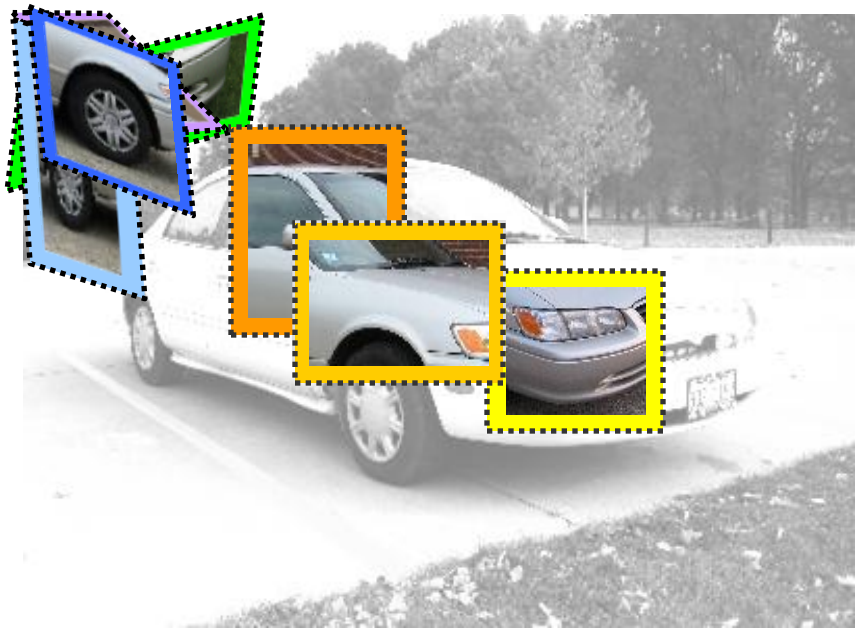
model

Algorithm

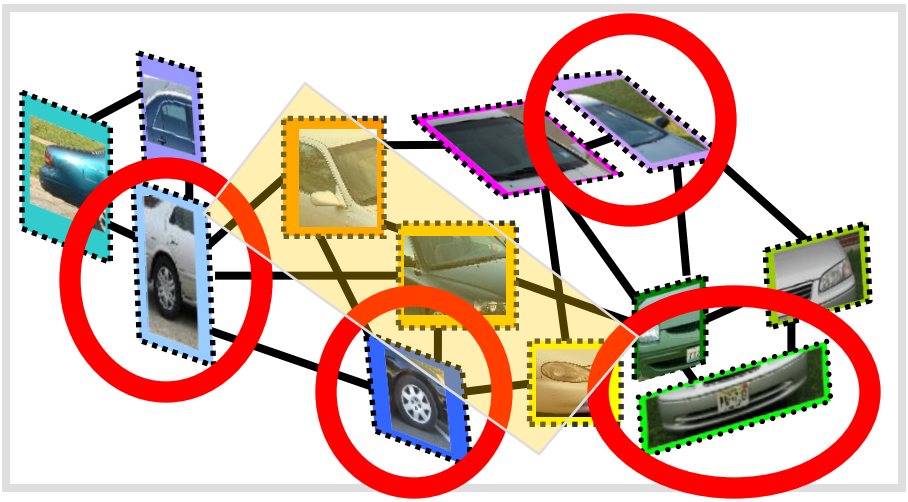
1. **Find hypotheses of canonical parts consistent with a given pose**
2. **Infer position and pose of other canonical parts**

Recognition

Query image



model

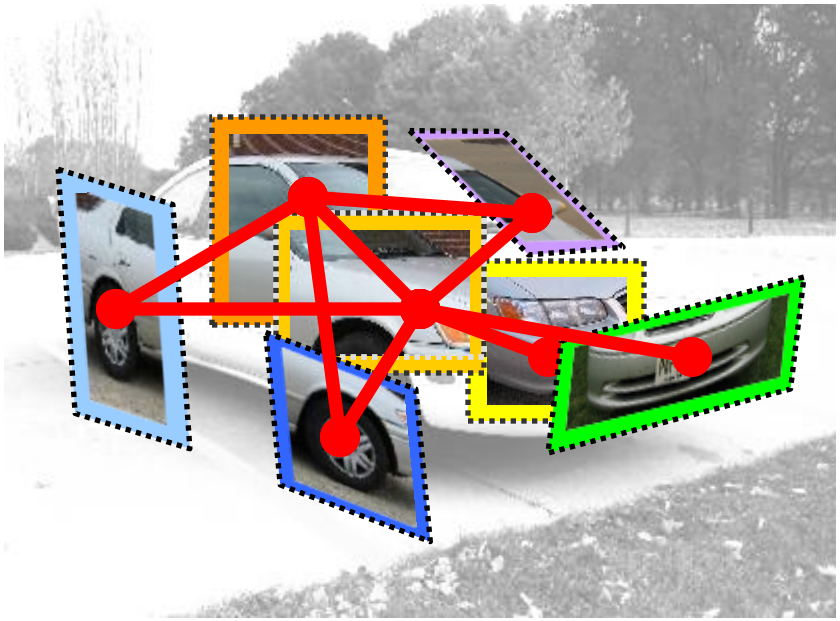


Algorithm

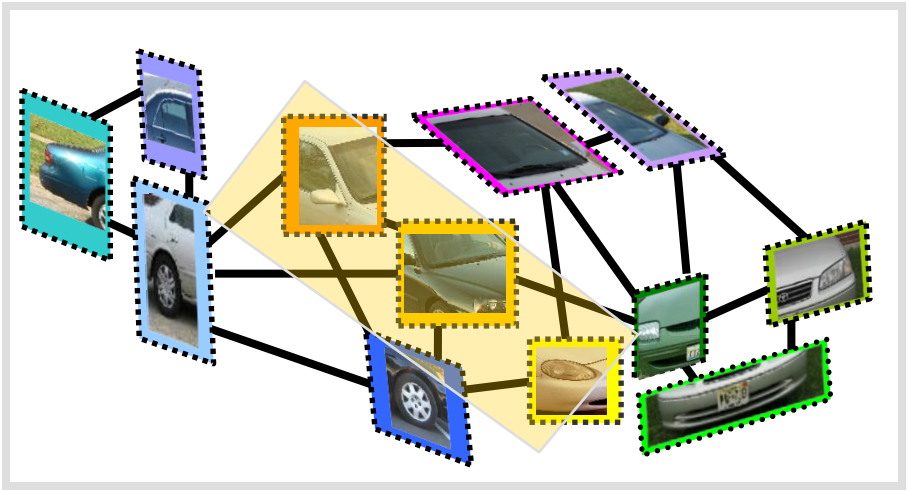
1. **Find hypotheses of canonical parts consistent with a given pose**
2. **Infer position and pose of other canonical parts**

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

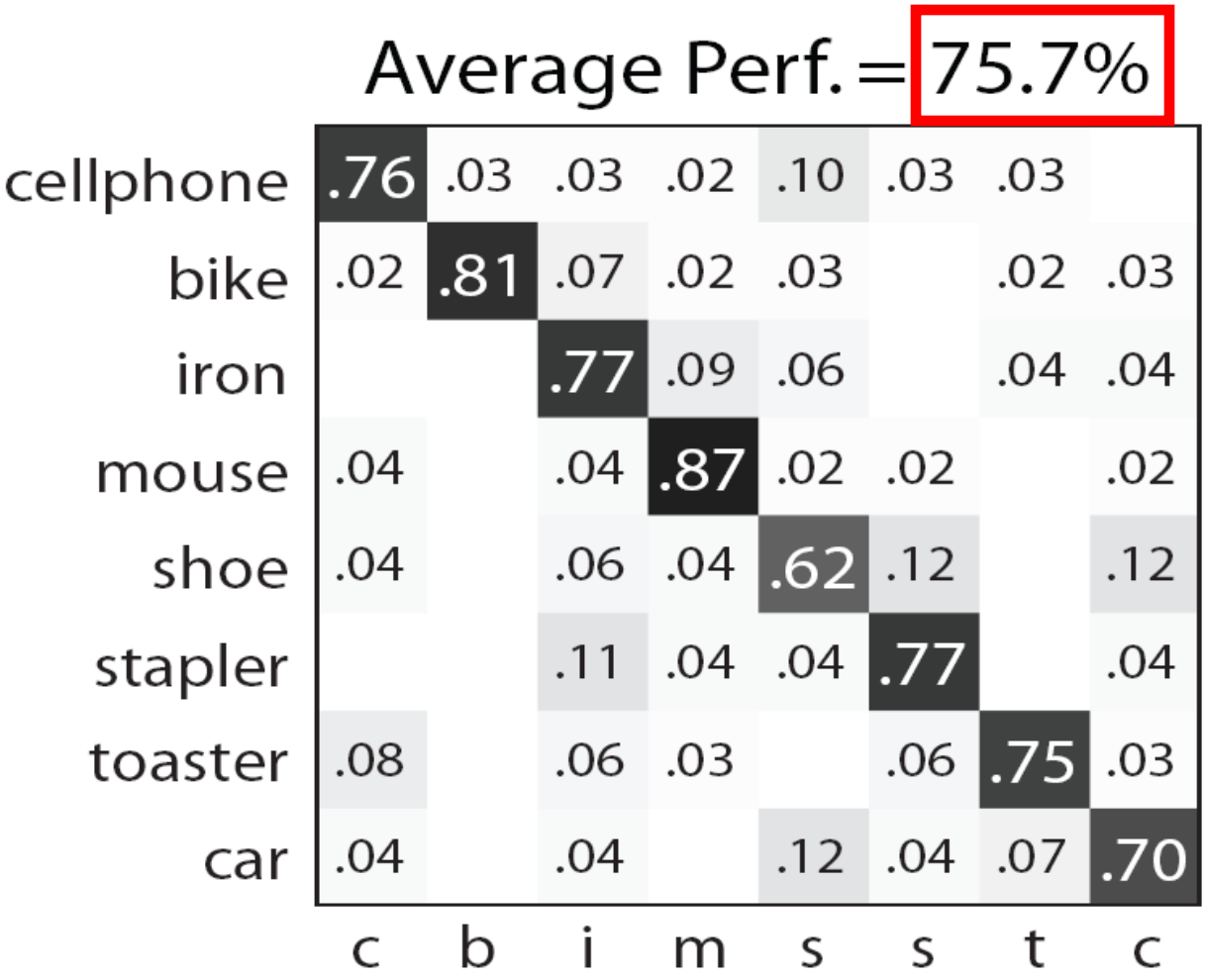
3. Optimize over A, G and s to find best combination of hypothesis

$$\text{Cost} = \sum_{h,k \in L} \sum_{i,j \in L} G(i,j,h,k)\, \delta_{ij} \delta_{hk} + \sum_{i,j \in L} A(i,j)\, \delta_{ij}$$

Examples

Category: iron
Azimuth = 135°
Zenith = 60°
Distance = medium

Classification accuracy



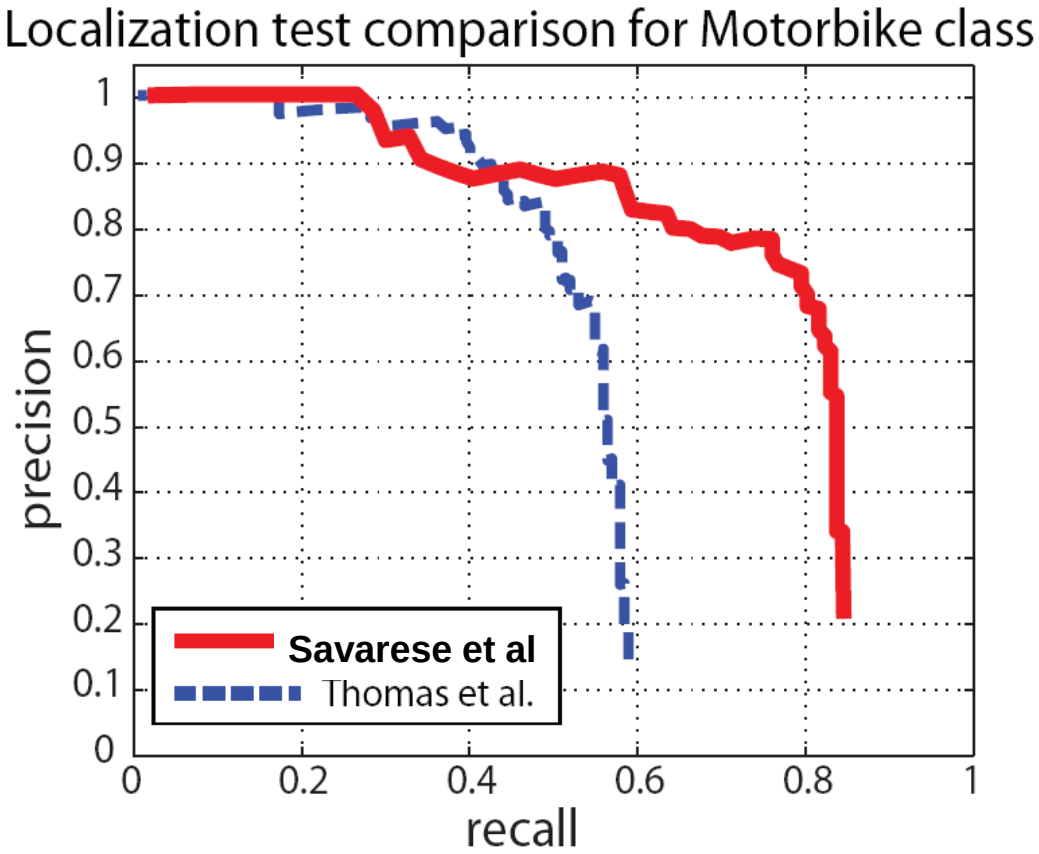
Failure example

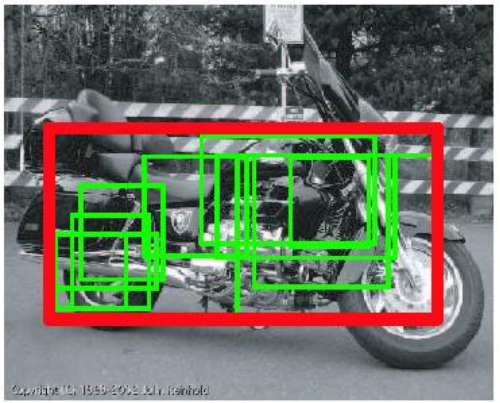
Category: shoe
Azimuth = 225°
Zenith = 30°
Distance = close

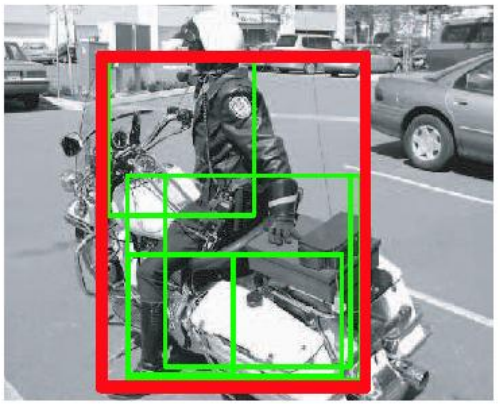
Linkage structure of canonical parts

Results

Savarese & Fei-Fei, '07, '08







Motorbike database from the PASCAL visual object classes challenge

Conclusion

	Single view	Thomas et al.	Kushal et al.	Savarese & Fei-Fei
View point invariant	X	√	√	√
No supervision	√	X category, instance, views	X → √ category, instance, Valid set	X → √ category
# Categories	~100	2	1	8
Share information across views	X	√	√	√
View synthesis	X	X	X	X → √
Sampling density	N.A.	X	X → √	X → √