Deformable Parts Model

an overview

Hossein Azizpour
CVAP - KTH

Outline

- Pictorial Structures and Deformable Parts Model
- A few extensions
- Our research line
- Sharing parts using Deformable Parts Model Framework

Pictorial Structure Felzenszwalb et al. IJCV05

INTRODUCTION

Pictorial Structure

- Energy Minimization
  - Appearance
  - Pairwise geometry

$$L^* = \arg\min_L \left(\sum_{i=1}^{n} m_i(l_i) + \sum_{(v_i,v_j)\in E} d_{ij}(l_i,l_j) \right)$$

- Simplify graph to tree
  - Dynamic programming

- Dynamic Programming
- Independence Assumption

$$\theta^* = \arg\max_\theta \prod_{k=1}^{m} p(I^k|L^k,\theta) \prod_{k=1}^{m} p(L^k|\theta).$$

- Gaussian estimation

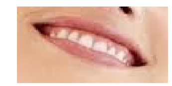
  of pair-wise part relations

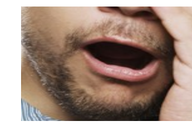
Part Based Models Felzenszwalb et al. PAMI10

- Mixture model
- Discriminative training
  - save for not considering denser graphs
- Weak-supervision (Latent variables)
  - Latent SVM
- Grammar based model
- Harvesting hard negative
- Post processing
  - Bounding box prediction
  - Context rescoring

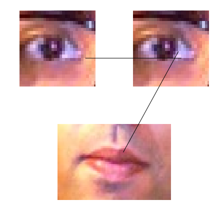
Object Detection Grammer Felzenszwalb et al. TR10

- 4 aspects of modeling
  - Parts (recursively)
    - e.g. Eyes, mouth, …
  - Sub-types
    - e.g. Smiling/Yawning eye

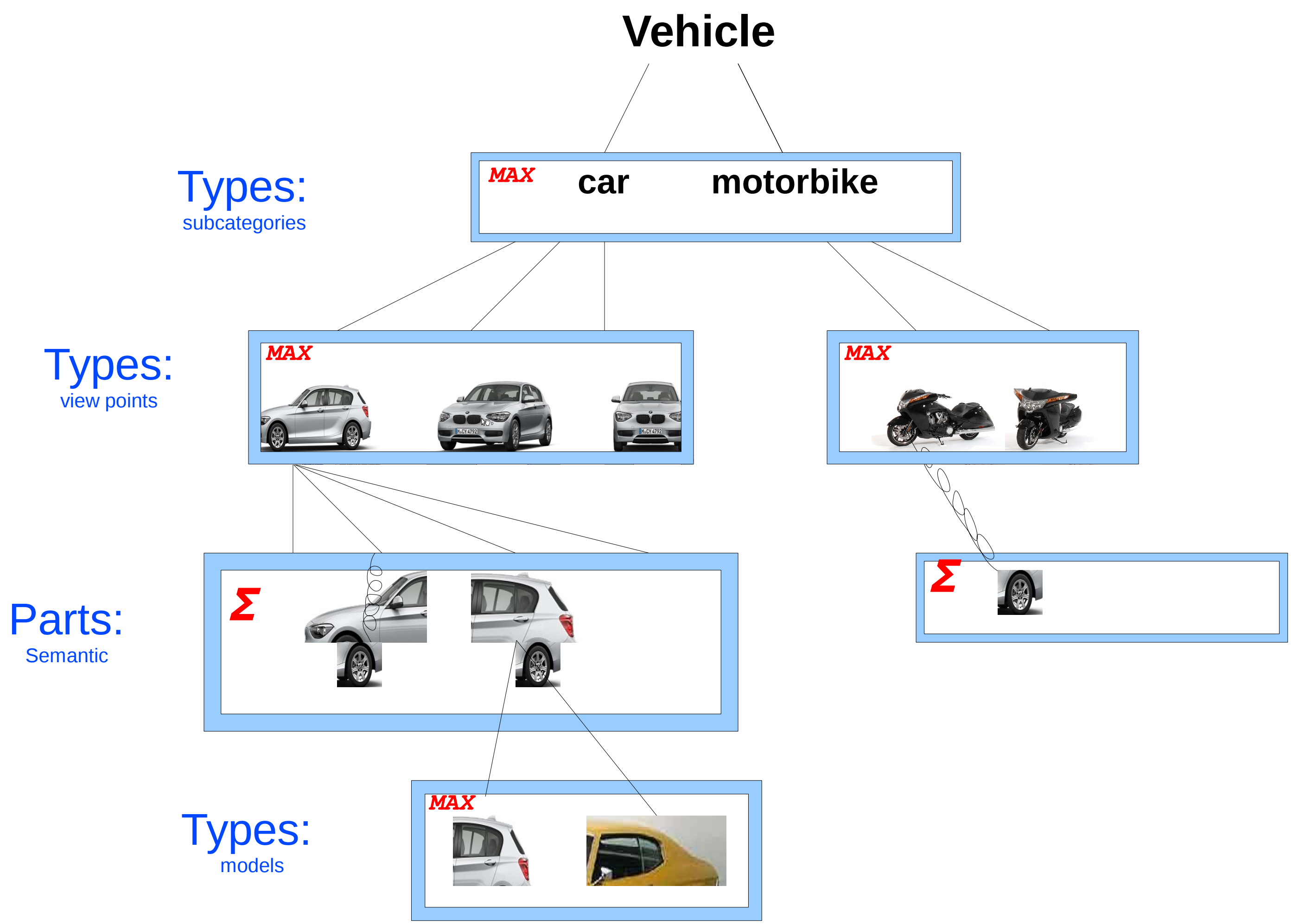




  - Geometric constellation



  - Appearance



And so on....

Object grammars

Model Formulation Felzenszwalb et all PAMI10

$$\text{score}(p_0, \ldots, p_n) =$$
$$\sum_{i=0}^{n} F_i' \cdot \phi(H, p_i) - \sum_{i=1}^{n} d_i \cdot \phi_d(dx_i, dy_i) + b,$$

$$\phi_d(dx, dy) = (dx, dy, dx^2, dy^2)$$

$$\beta = (F_0', \ldots, F_n', d_1, \ldots, d_n, b).$$
$$\psi(H, z) = (\phi(H, p_0), \ldots \phi(H, p_n),$$
$$-\phi_d(dx_1, dy_1), \ldots, -\phi_d(dx_n, dy_n), 1)$$

Learning (LSVM) Felzenszwalb et all PAMI10

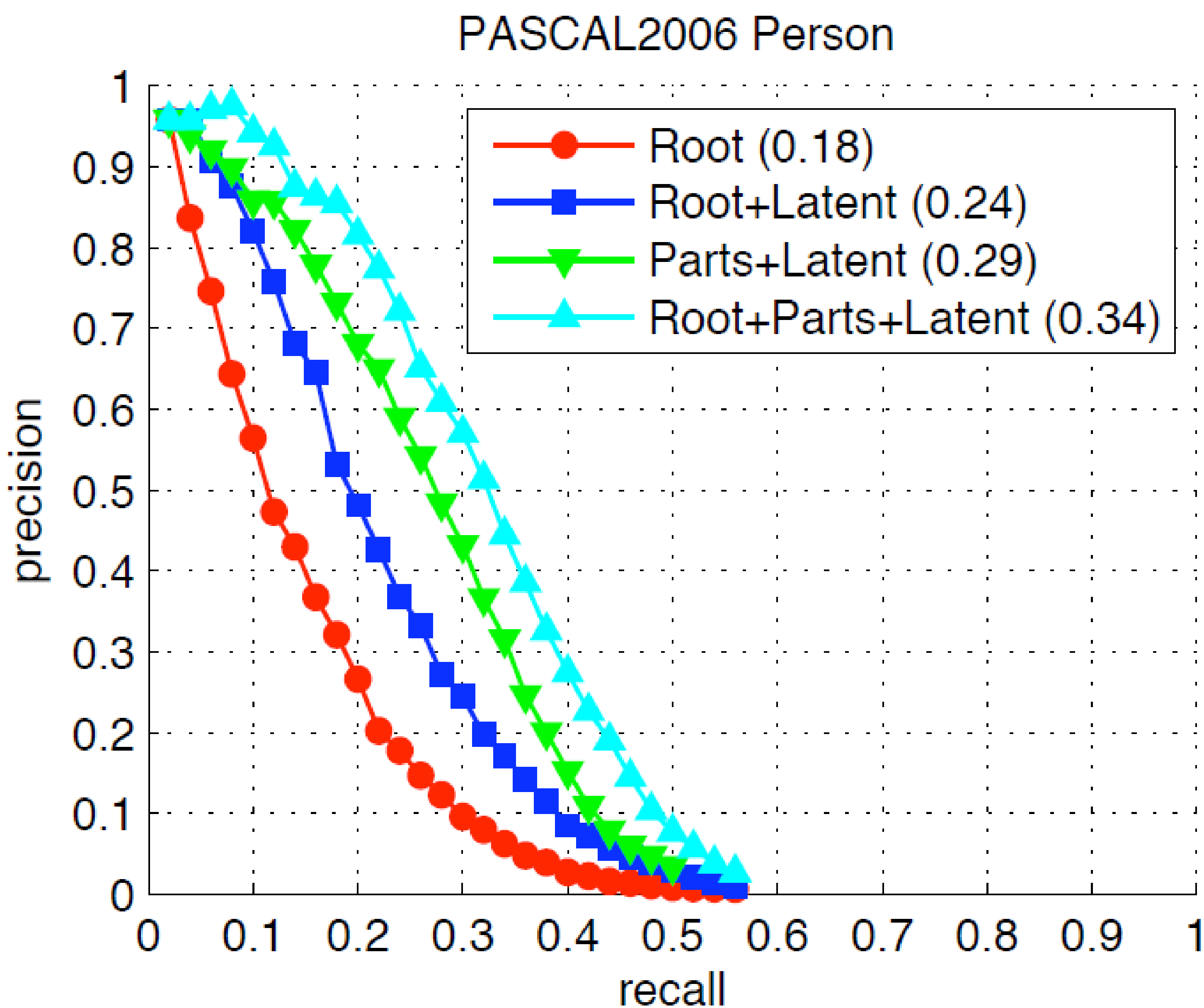
- SVM Optimization

$$L_D(\beta) = \frac{1}{2}||\beta||^2 + C\sum_{i=1}^{n} \max(0, 1 - y_i f_\beta(x_i)),$$

  - Linear SVM $f_\beta(x) = \beta \cdot \Phi(x)$
  - Latent SVM $f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z).$
- Semi-Convexity
- Coordinate-Descent

1) *Relabel positive examples:* Optimize $L_D(\beta, Z_p)$ over Z_p by selecting the highest scoring latent value for each positive example,
$z_i = \text{argmax}_{z \in Z(x_i)} \beta \cdot \Phi(x_i, z).$
2) *Optimize beta:* Optimize $L_D(\beta, Z_p)$ over β by solving the convex optimization problem defined by $L_{D(Z_p)}(\beta)$.

Results



PASCAL2006 Person
Root (0.18)
Root+Latent (0.24)
Parts+Latent (0.29)
Root+Parts+Latent (0.34)
precision
recall
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Cascade Object Detection with Deformable Part Models

Felzenszwalb et al. CVPR10

- 10 times Faster
- Same performance
- Simplified Part Appearance Model
- General Grammar Model

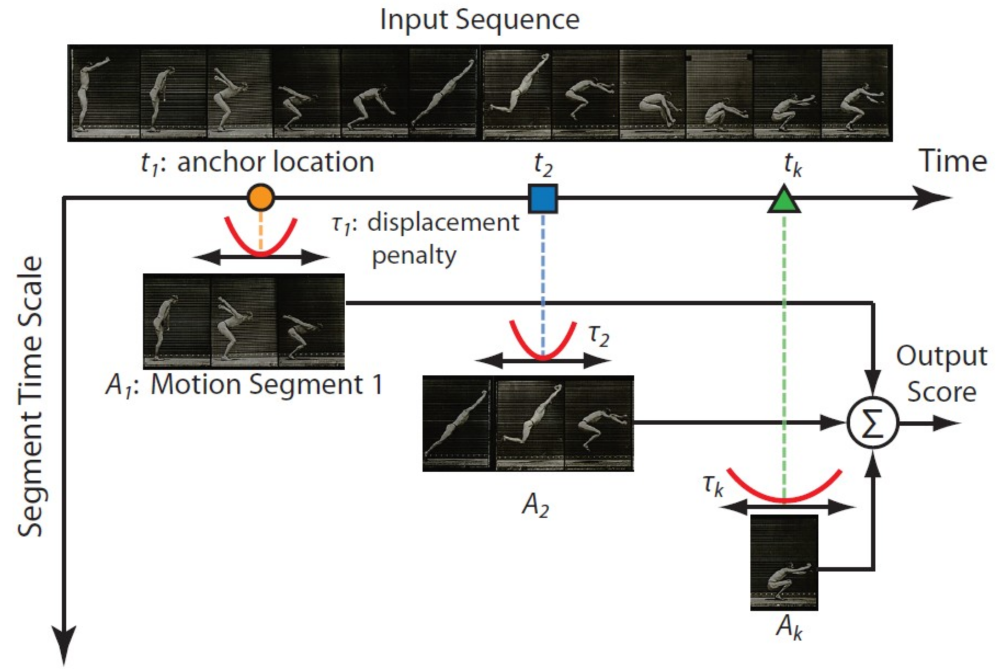
	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	inria
Speedup factor	22.7	22.1	16.5	11.6	22.1	36.0	13.3	25.6	23.4	23.2	29.8	15.2	16.2	32.6	12.7	23.3	32.8	18.1	23.3	27.2	13.5
Baseline AP	21.1	43.1	10.6	12.2	24.0	42.2	48.0	15.9	13.4	19.0	7.1	10.7	31.3	32.9	34.4	12.0	20.3	20.8	29.3	36.3	80.1
Cascade AP	21.1	42.9	10.4	12.4	24.1	42.5	48.1	15.5	13.4	19.0	8.0	10.7	31.3	33.0	34.8	12.0	20.3	20.2	28.8	36.5	80.1

Modeling Temporal Structure of Decomposable Motion Segments for Activity Classification

Juan Carlos Niebles, Chih-Wei Chen, Li Fei-Fei ECCV10

- Complex actions: temporal composition of simple action
- BOF X^2 RBF
- KTH

Algorithm	Perf.
Ours	91.3%
Wang et al. [28]	92.1%
Laptev et al. [5]	91.8%
Wong et al. [8]	86.7%
Schuldt et al. [27]	71.5%
Kim et al. [29]	95%



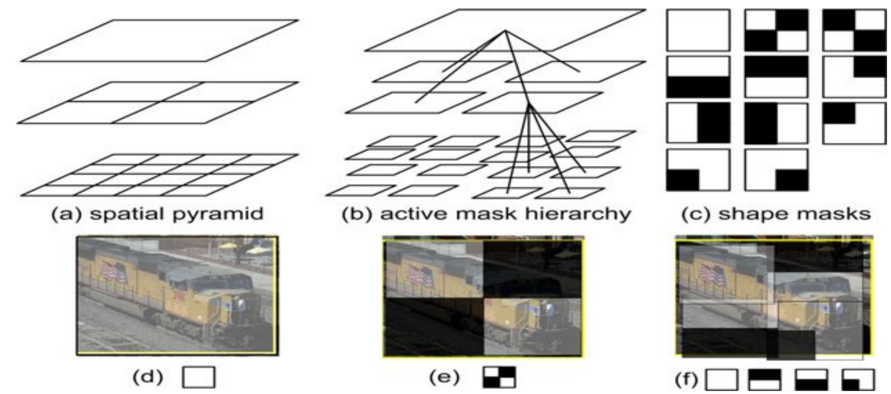
Olympic dataset

high-jump	**68.9%**	52.4%	javelin-throw	**74.6%**	61.1%
long-jump	**74.8%**	66.8%	hammer-throw	**77.5%**	65.1%
triple-jump	**52.3%**	36.1%	discus-throw	**58.5%**	37.4%
pole-vault	**82.0%**	47.8%	diving-platform	87.2%	**91.5%**
gymnastics-vault	86.1%	**88.6%**	diving-springboard	77.2%	**80.7%**
shot-put	**62.1%**	56.2%	basketball-layup	**77.9%**	75.8%
snatch	**69.2%**	41.8%	bowling	**72.7%**	66.7%
clean-jerk	**84.1%**	83.2%	tennis-serve	**49.1%**	39.6%
			Average (AAP)	**72.1%**	62.0%

Active Mask Hierarchies for Object Detection

Yuanhao Chen, Long Zhu, Alan Yuille, William Freeman CVPR10 and ECCV10

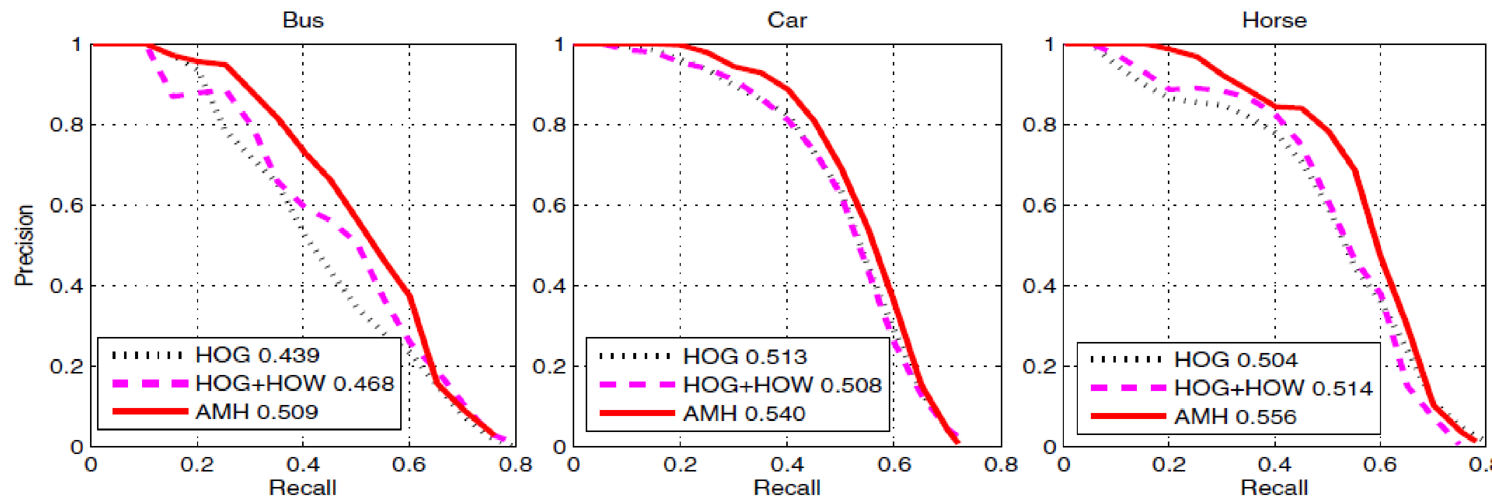
- HOG + BOF
- Shape Masks
- iCCP
- MKL + X^2
- Spatial Pyramid + Parts based models

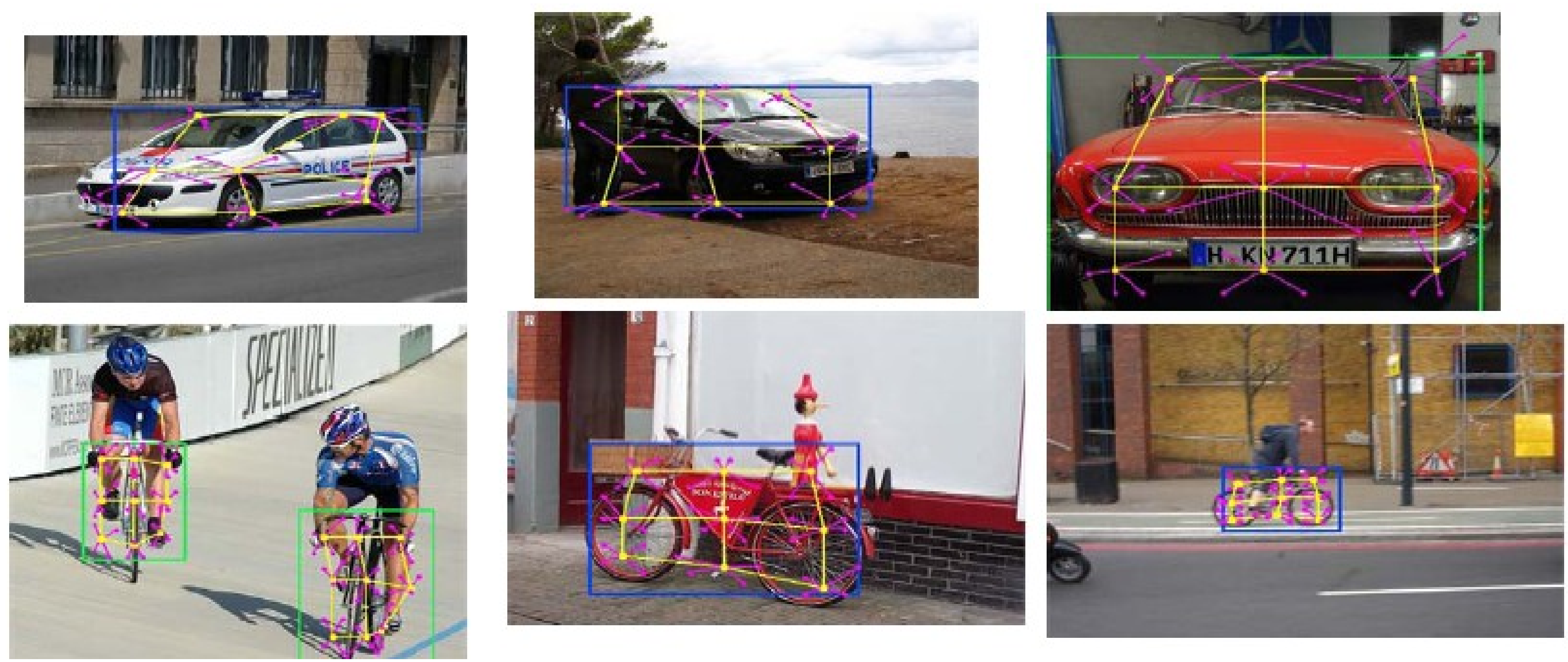


class	Ave.	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow
Active Mask Hierarchies (AMH)	**.338**	.348	.544	**.155**	.146	.244	**.509**	**.540**	**.335**	**.206**	.228
Hierarchy without masks [8]	.296	.294	.558	.094	.143	**.286**	.440	.513	.213	.200	.193
UoCTTI-1 (Part-based) [2]	.268	.290	.546	.006	.134	.262	.394	.464	.161	.163	.165
UoCTTI-2 (Part-based) [2]	.298	.328	**.568**	.025	**.168**	.285	.397	.516	.213	.179	.185
MKL-1 (Pyramid-based) [1]	.292	.366	.425	.128	.145	.151	.464	.459	.255	.144	.304
MKL-2 (Pyramid-based) [1]	.321	**.376**	.478	.153	.153	.219	.507	.506	.300	.173	**.330**
V07 [9]	—	.262	.409	.098	.094	.214	.393	.432	.240	.128	.140

	Ave.	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
Active Mask Hierarchies	**.338**	**.344**	**.241**	**.556**	**.473**	.349	**.181**	.202	**.303**	.413	.433
Hierarchy without masks [8]	.296	.252	.125	.504	.384	.366	.151	.197	.251	.368	.393
UoCTTI-1 (Part-based) [2]	.268	.245	.050	.436	.378	.350	.088	.173	.216	.340	.390
UoCTTI-2 (Part-based) [2]	.298	.259	.088	.492	.412	**.368**	.146	.162	.244	.392	.391
MKL-1 (Pyramid-based) [1]	.292	.190	.160	.490	.460	.215	.110	**.245**	.264	.426	.408
MKL-2 (Pyramid-based) [1]	.321	.225	.215	.512	.455	.233	.124	.239	.285	**.453**	**.485**
V07 [9]	—	.098	.162	.335	.375	.221	.120	.175	.147	.334	.289

Active Mask Hierarchies for Object Detection Yuanhao Chen, Long Zhu, Alan Yuille, William Freeman CVPR10 and ECCV10

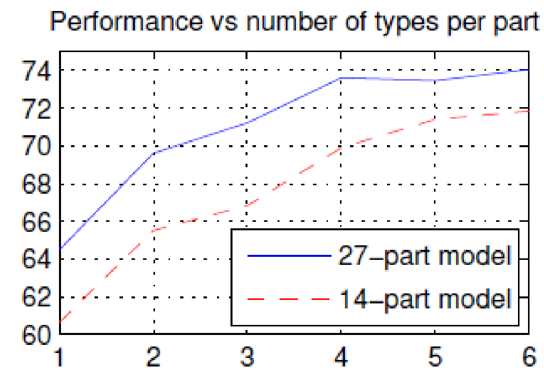
- 



Articulated Pose Estimation using Flexible Mixtures of Parts

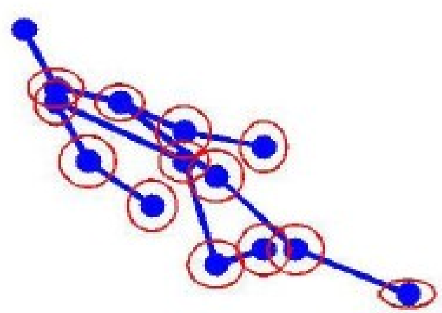
Y. Yang, D. Ramanan CVPR11

- Articulated pose estimation
- Sub-types to model part orientations
- Extension of formulation
  - Prior score for different subtypes
  - Model corelation of parent/child subtypes



$$S(t) = \sum_{i \in V} b_i^{t_i} + \sum_{ij \in E} b_{ij}^{t_i,t_j}$$

$$S(I,p,t) = S(t) + \sum_{i \in V} w_i^{t_i} \cdot \phi(I,p_i) + \sum_{ij \in E} w_{ij}^{t_i,t_j} \cdot \psi(p_i - p_j) \quad (2)$$



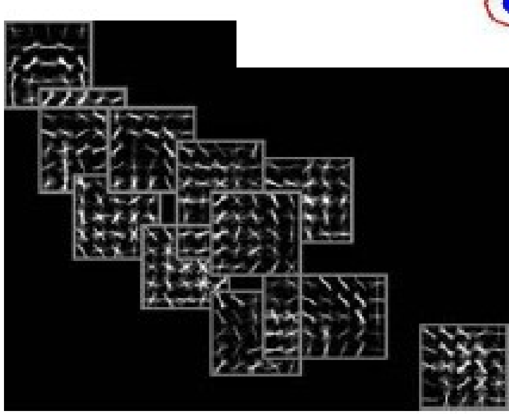


Image Parse Testset

Method	Torso	Head	Upper legs	Lower legs	Upper arms	Lower arms	Total
R [23]	52.1	37.5	31.0	29.0	17.5	13.6	27.2
ARS [1]	81.4	75.6	63.2	55.1	47.6	31.7	55.2
JEa [15]	77.6	68.8	61.5	54.9	53.2	39.3	56.4
SNH [29]	91.2	76.6	71.5	64.9	50.0	34.2	60.9
JEb [14]	85.4	76.1	73.4	65.4	64.7	46.9	66.2
Our Model	97.6	93.2	83.9	75.1	72.0	48.3	74.9

Shared Parts for Deformable Part-based Models

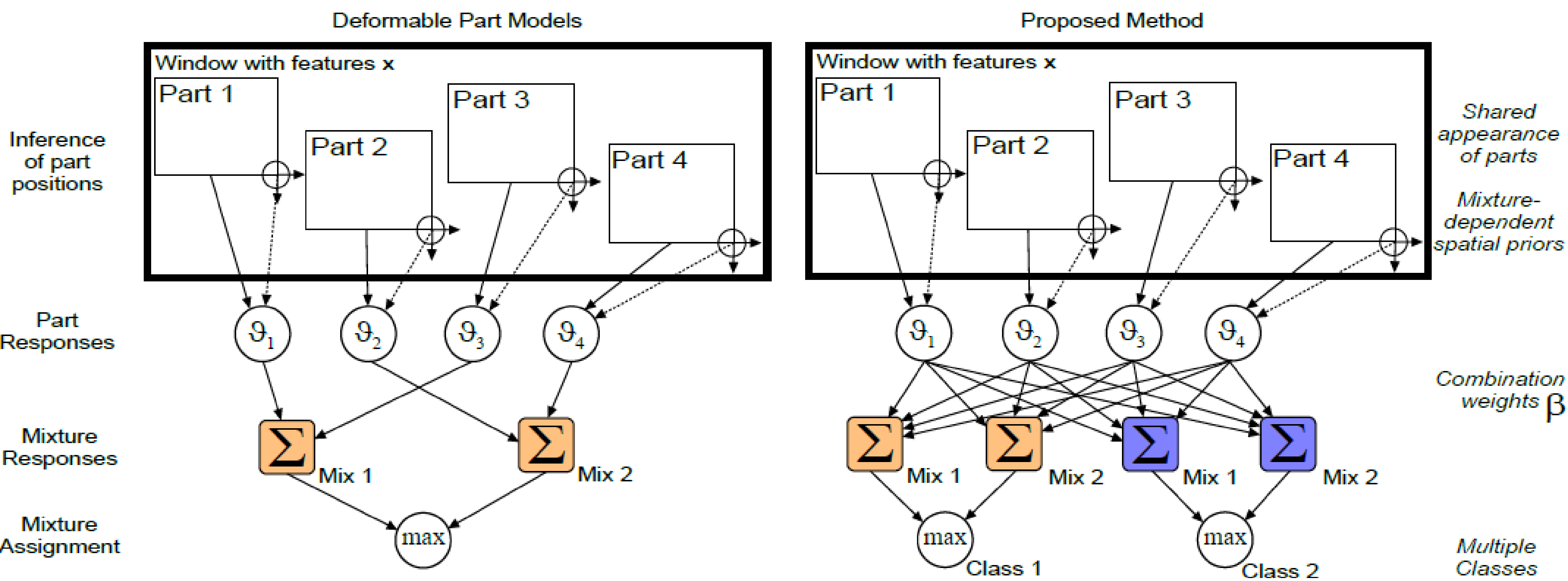
- Builds on Felzenswalb Deformable Parts Model
- Motivation: problems with adding more mixture components and parts to model
  - Increase in number of parameters and decrease in number of training per component -> poor generalization
  - Necessary resources increase linearly

Results

- Share parts between mixtures and classes
  - Increasing training examples available to each class by sharing
  - Less parameters -> more compact models
  - Less expensive training and testing

Model – Share parts within a class

- Natural to have the same appearance over different viewpoints
- Intermediate visual modes can be captured



Model – Share parts within a class

- Appearance shared among components
- Spatial configuration can vary
- Different weights in different mixtures
  - Decides whether a part is visible in a component
  - Decides how discriminative is each part (relative to others) contributing to a component

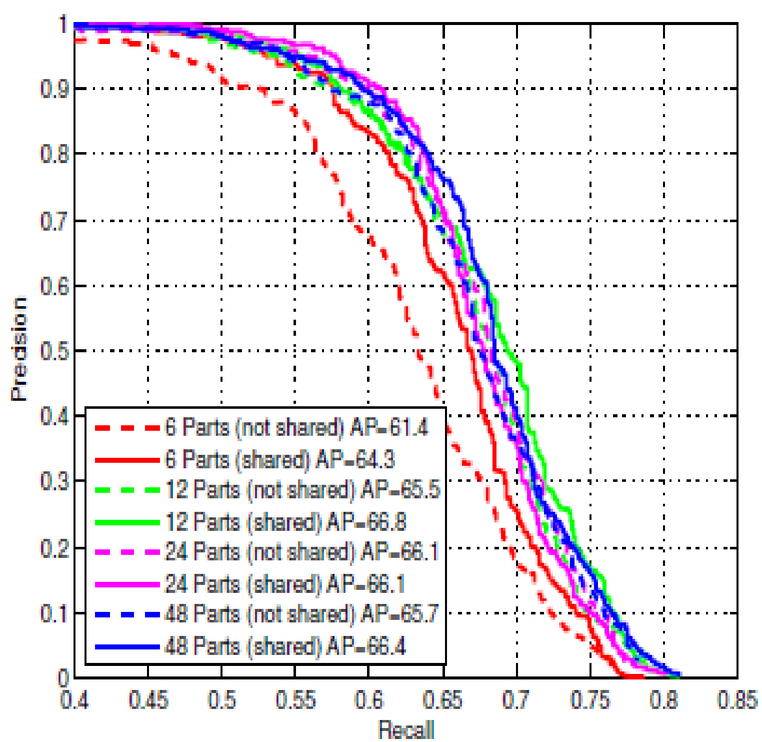
$$g^l(x) = b_l + \sum_{i=1}^{p} \beta_i^l \vartheta\left(x, w^i, v^{l,i}, a^{l,i}\right)$$

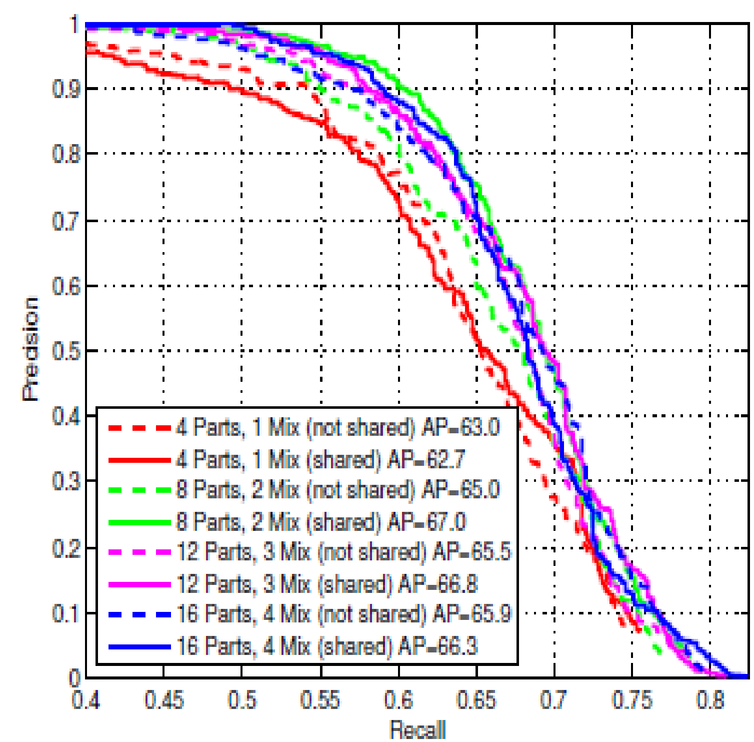
Results

-

	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow
not shared	**26.7**	34.9	**0.2**	0.5	0.1	32.5	32.4	5.1	**4.2**	**6.5**
shared	24.7	**38.2**	0.0	**1.2**	**0.2**	**33.3**	**37.7**	**7.3**	1.4	4.6

table	dog	horse	mbike	pers	plant	sheep	sofa	train	tv	mean
4.7	5.5	15.9	29.7	10.3	0.0	**17.3**	4.1	**9.7**	**24.4**	13.2
8.1	**8.1**	**21.5**	**31.8**	**11.5**	**6.3**	17.0	**5.1**	9.6	23.9	**14.6**





Thank You!