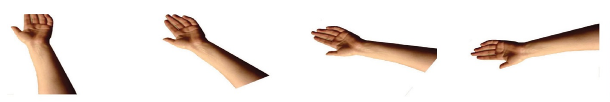
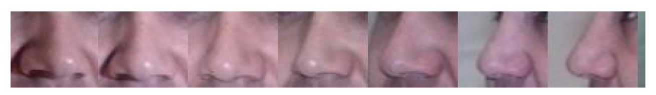
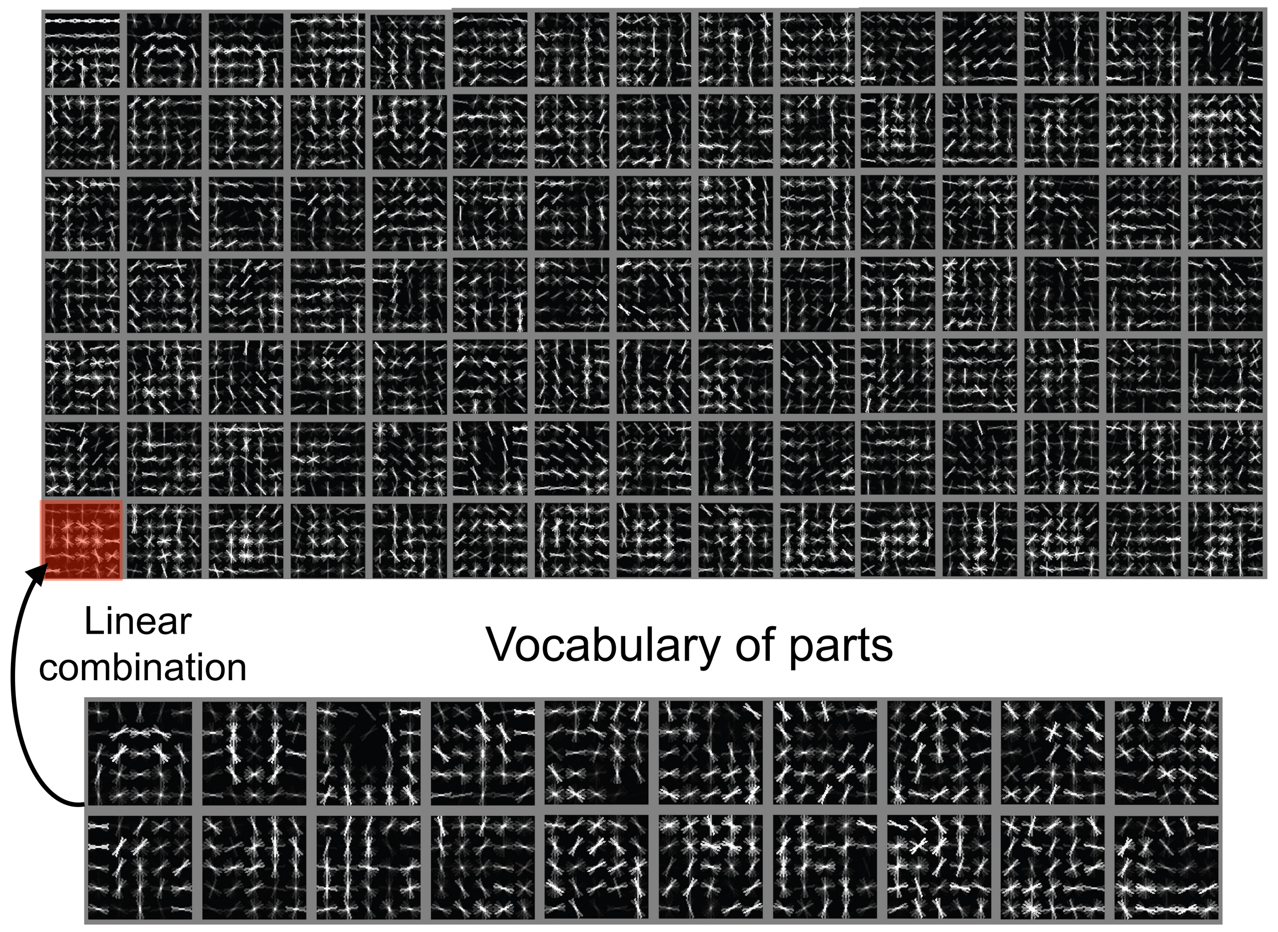
Steerable Part Models

Hamed Pirsiavash, Deva Ramanan   Computational Vision Lab, University of California Irvine, CA, USA   {hpirsiav, dramanan} @ ics.uci.edu



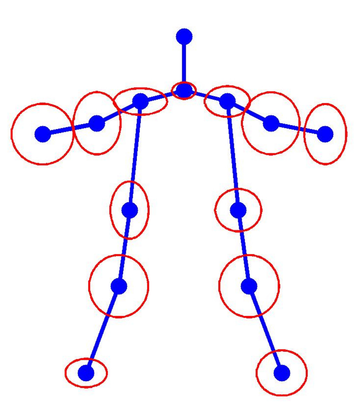


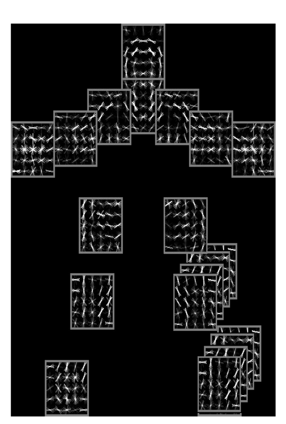
In-plane rotation   Out-of-plane rotation

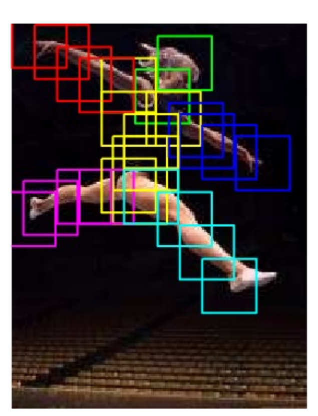


Steerable basis

Background on Part Models







Appearance feature eg, HOG

$$\underbrace{\text{score}(I,l,t)}_{\text{Score of this placement}} = \Big[\sum_i^m \underbrace{w_i \cdot \phi_a(I,l_i)}_{\text{Score for the } i\text{'th filter}}\Big] + \underbrace{w_s \cdot \phi_s(l,t)}_{\text{Score for all springs}}$$

Motivation

Large variation in appearance:
Change in view point, deformation, and scale

First solution:
Introduce mixtures → Discretely handle appearance variation

What about a large number of mixtures?

- Not scalable to a large part vocabulary
- Over-fitting due to high dimensional learning problem

Approach:

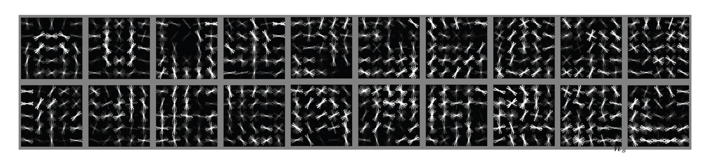
(1) Learn low-dimensional filter banks, not high-dimensional parameter vectors
(2) Represent large vocabulary of parts with a small set of separable basis filters

Inspired by steerable filters in image processing

Citation: Manduchi, Perona, Shy,
"Efficient Deformable Filter Banks", IEEE Trans Signal Proc. 1998

$$w_i = \sum_{j=1}^{n_s} s_{ij} b_j$$

s_{ij}: Steering coefficient; b_j → Set of basis filters



Can be written as a rank restriction on filter bank of parameters
Citation: Pirsiavash, Ramanan, Fowlkes,
"Bilinear Classifiers for Visual Recognition", NIPS 2009

$$\underbrace{\Big[\quad W \quad\Big]}_{\text{Size of part vocabulary}} = \underbrace{\Big[\, B \,\Big]}_{n_s}\Big[\quad S^T \quad\Big]$$

Size of part vocabulary n_s : Number of basis filters

Learning: Structured SVM

Eq (1)

$$L(w) = \frac{1}{2}w^T w + C\sum_n \max_{z \in Z_n}[0, 1 - y_n w^T \phi(I_n, z)]$$

$$Z_n = \{z_n\} \quad \forall n \quad \text{s.t.} \quad y_n = 1$$

$$Z_n = \{\text{unrestricted}\} \quad \forall n \quad \text{s.t.} \quad y_n = -1$$

Coordinate decent algorithm: repeat

1. Fix basis, learn coefficients

$$(S^*, w_s^*) = \text{argmin}_{S, w_s} L(B^*, S, w_s)$$

2. Fix coefficients, learn basis

$$(B^*, w_s^*) = \text{argmin}_{B, w_s} L(B, S^*, w_s)$$

Convex steps
→ Each step can be written as Eq (1) after change of basis.

Steerability and Separability

b_j itself is a matrix → write it in separable form

$$B_j = \sum_{k=1}^{n_k} c_{jk} f_{jk}^T$$

Share the sub-space by forcing $f_{jk} = f_k$

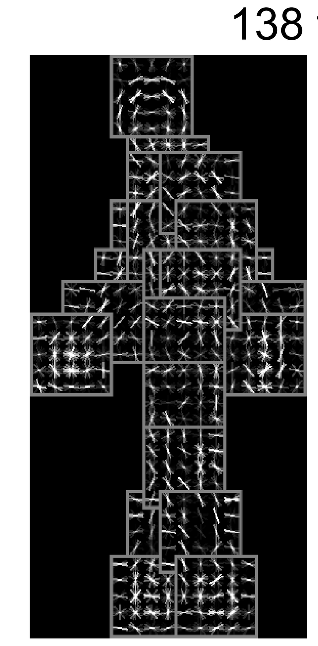
n_k : Number of dimensions of subspace

Experiments

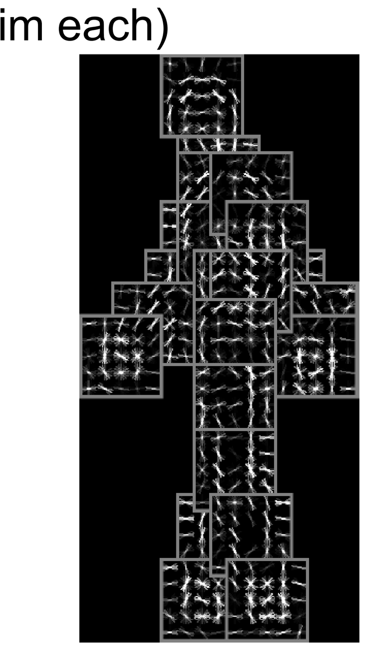
Human pose estimation   Face pose estimation   Object detection

Human pose estimation

138 filters (800 dim each)

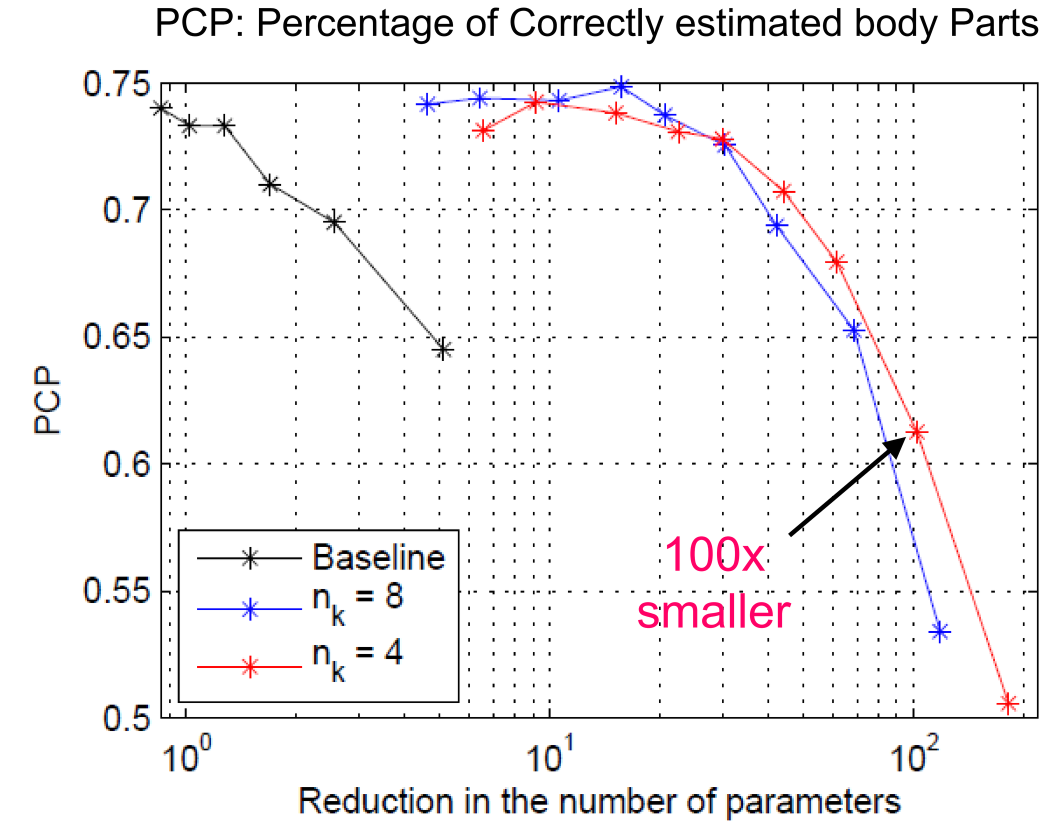


Original model
Yang, Ramanan, CVPR' 11

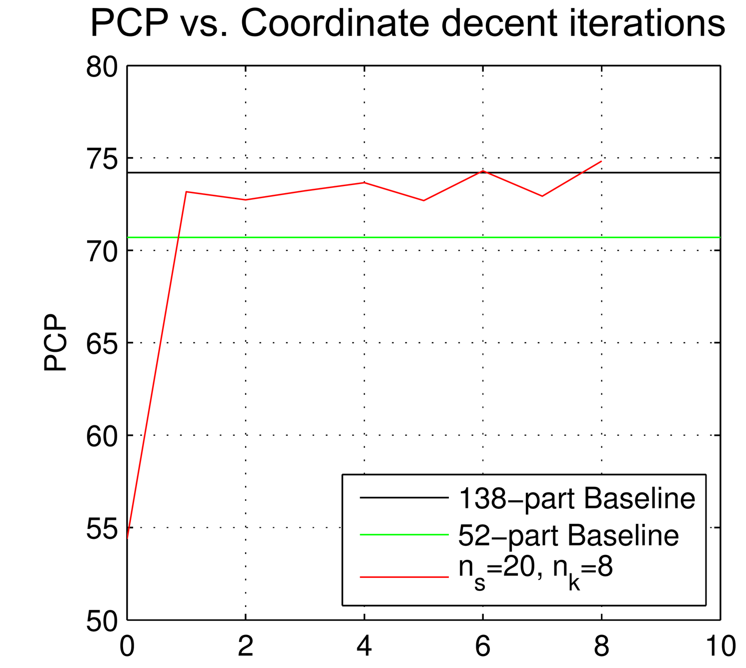


Reconstructed model
(15x smaller)
$n_s = 20, \ n_k = 8$

Reduction in the model size
PCP: Percentage of Correctly estimated body Parts

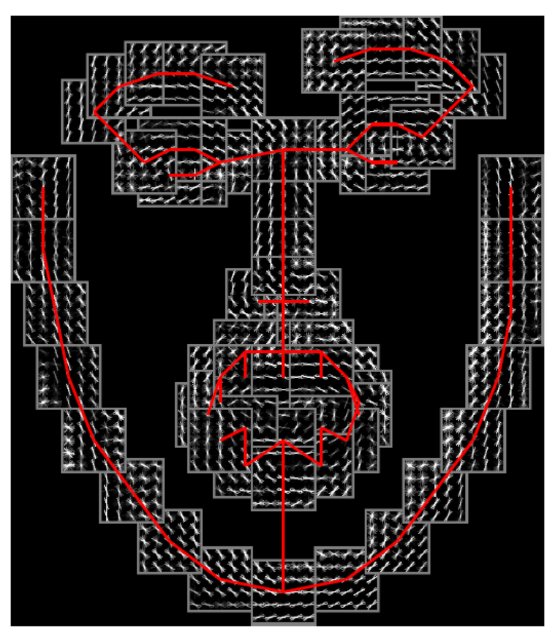


PCP vs. Coordinate decent iterations

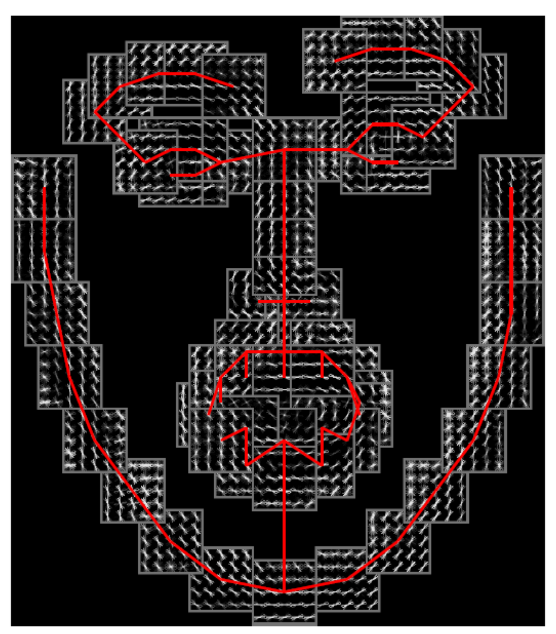


Face detection, pose estimation, and landmark localization

1050 filters (800 dim each)



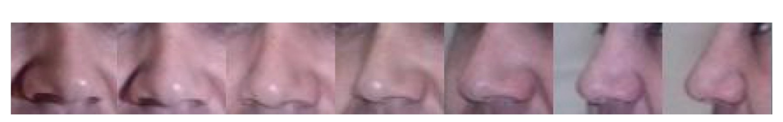
Original model
Zhu, Ramanan, CVPR' 12



Reconstructed model
(24x smaller)

Method	Reduction in # params	# basis n_s	Subspace dim n_k	Accuracy of pose estimation	Localization error (mse)
1050-part baseline	1	-	-	91.4	0.0234
99-part shared baseline	10.6	-	-	81.5	0.0281
Our Model	7.2	93	8	91.6	0.0236
Our Model	24.3	30	4	89.9	0.0256

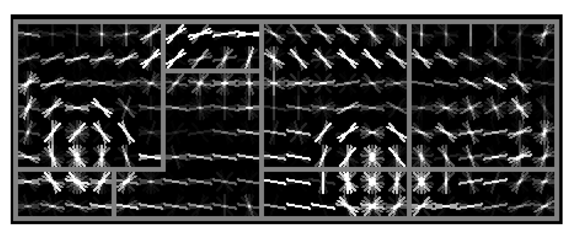
Our model outperforms manually defined "hard-sharing": only one part for all views of nose



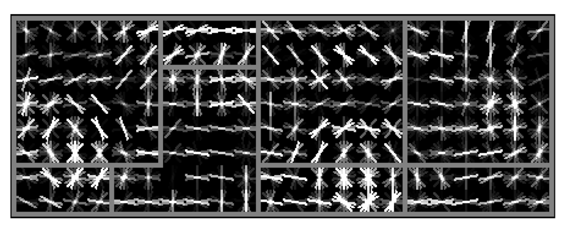
PASCAL object detection

20 categories, 480 filter, (800 dim each)

Share basis across categories
Soft sharing: a "wheel" template can be shared between "car" and "bike" categories



Original model
Felzenszwalb, Girshick, Mc-Allester, Ramanan, TPAMI' 10



Reconstructed model
(3x smaller)

Conclusion

- We write part templates as linear filter banks.
- We leverage existing SVM-solvers to learn steerable representations using rank-constraints.
- We demonstrate impressive results on three diverse problems showing improvements up to 10x-100x in size and speed.
- We demonstrate that steerable structure can be shared across different object categories.