Steerable Part Models

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

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

\[
W = [B] [ST] 
\]

Size of part vocabulary \(N_B\) : Number of basis filters

Learning: Structured SVM

\[
L(w) = \frac{1}{2} w^T w + C \sum_{n} \max[0, 1 - y_n w^T \phi(I_n, z)]
\]

\(Z_n = \{x_n\} \cap \text{s.t.} \ y_n = 1\)

\(Z_n = \{\text{unrestricted}\} \cap \text{s.t.} \ y_n = -1\)

Coordinate decent algorithm: repeat
1. Fix basis, learn coefficients
   \(S^*, w^*_n = \text{argmin}_{S, w_n} L(B^*, S, w_n)\)

2. Fix coefficients, learn basis
   \(B^*, w^*_n = \text{argmin}_{B, w_n} L(B, S^*, w_n)\)

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
138 filters (800 dim each)

Face detection, pose estimation, and landmark localization
1050 filters (800 dim each)

Face detection, pose estimation, and landmark localization
20 categories, 480 filter, (800 dim each)

PASCAL object detection
20 categories, 480 filter, (800 dim each)

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.