Bilinear Classifiers for Visual Recognition

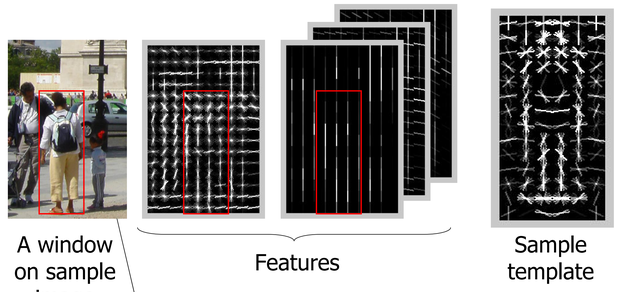
Hamed Pirsiavash, Deva Ramanan, Charless Fowlkes Computational Vision Lab, University of California Irvine, CA USA

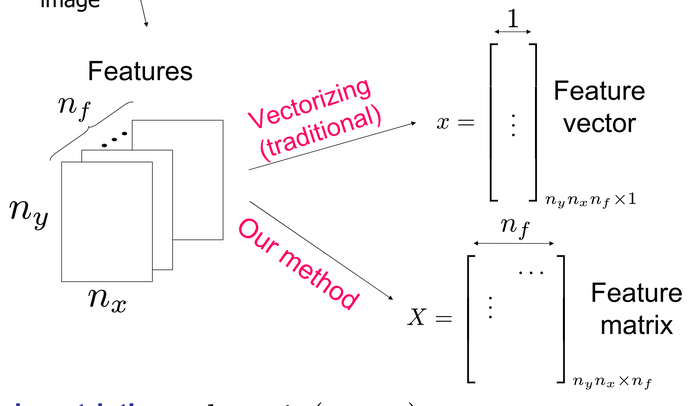
Main ideas:

-Treat image features and templates as matrices or tensors rather than simply “vectorizing” them

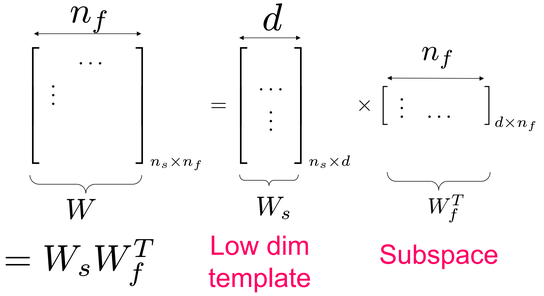
-Learn a low rank template by performing alternating minimization using standard linear SVM solver

-Share a subset of parameters (subspace) across different problems (transfer learning)





Rank restriction: $d \leq \min(n_s, n_f)$ where $n_s = n_y n_x$



$W = W_s W_f^T$ Low dim template Subspace

Linear SVM: For a given set of training pairs: $\{X_n, y_n\}$

$$\min_W L(W) = \underbrace{\tfrac{1}{2} Tr(W^T W)}_{\text{Regularizer}} + C \underbrace{\textstyle\sum_n \max(0, 1 - y_n Tr(W^T X_n))}_{\text{Constraints}}$$

Detection function:

Was Linear: $f(X) = Tr(W^T X) > 0$

Now it is Bilinear: $f(X) = Tr(W_s^T X W_f) > 0$

Learning: It is *biconvex* so we use coordinate descent with off-the-shelf linear SVM solver in the loop

Change of basis: $\tilde{W}_f = A^{\frac{1}{2}} W_f^T,\ \tilde{X}_n = A^{-\frac{1}{2}} W_s^T X_n,\ A = W_s^T W_s$

$$\min_{\tilde{W}_f} L(\tilde{W}_f, W_s) = \tfrac{1}{2} Tr(\tilde{W}_f^T \tilde{W}_f) + C \textstyle\sum_n \max(0, 1 - y_n Tr(\tilde{W}_f^T \tilde{X}_n))$$

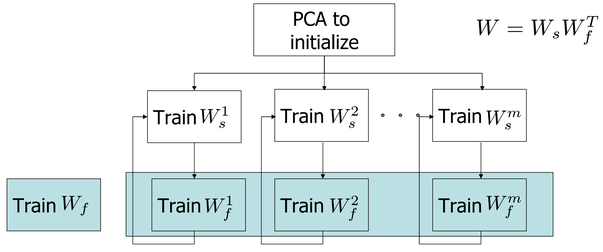
Advantages :

-Rank constraint provides natural *regularization*.

-10X runtime *speedup* with no loss in performance.
(Fewer number of convolutions in detection)

-Learn shared subspaces across multiple classes or training sets to allow for *transfer learning*.
(e.g., share subspace between “human detector” and “cat detector”)

Joint minimization: $\min_{\tilde{W}_f} \sum_{i=1}^{m} L(\tilde{W}_f, W_s^i)$



Related work :

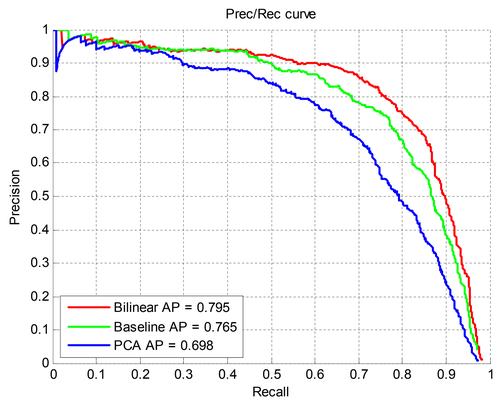
-Tenenbaum & Freeman’s *bilinear model* (Neural Computation’00)

-Wolf et al’s *low-rank SVM* paper (CVPR’07)

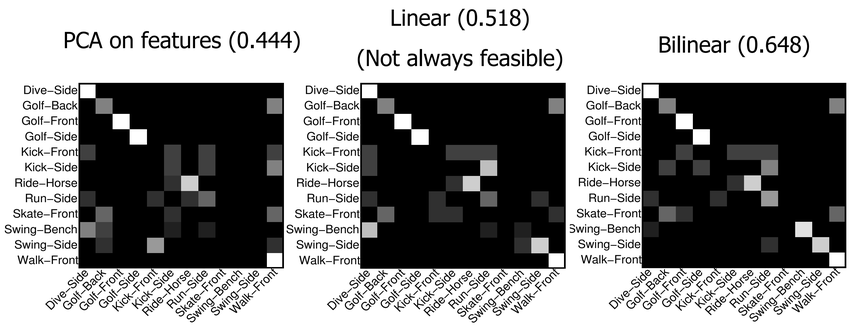
-Ando and Zhang's *transfer learning* paper (JMLR’05)

Experiment 1: Pedestrian detection on INRIA motion dataset

$$n_s = 14 \times 6,\ n_f = 84,\ d = 5$$



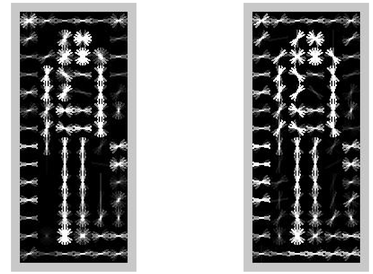
Experiment 2: Human action recognition on UCF sports dataset



Experiment 3: Sharing subspace on UCF sports dataset

(Using only two training samples for each class)

Iteration 1 Refined at Iteration 2



Average classification rate

	Coordinate decent iteration 1	Coordinate decent iteration 2
Independent bilinear (C=0.01)	0.222	0.289
Joint bilinear (C=0.1)	0.269	0.356