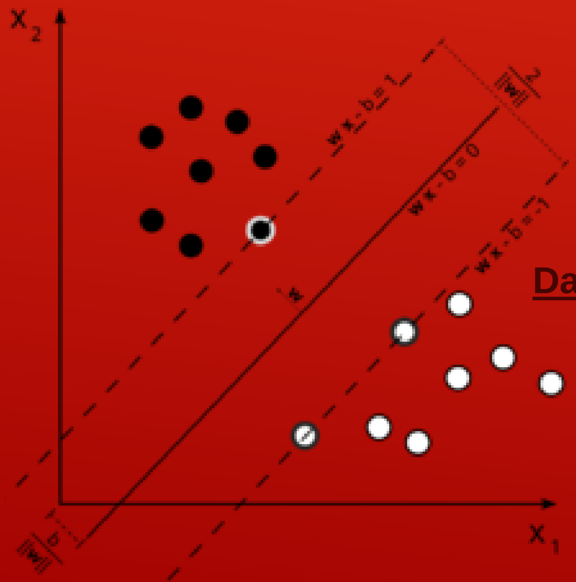
DISCRIMINATIVE DECORELATION FOR CLUSTERING AND CLASSIFICATION

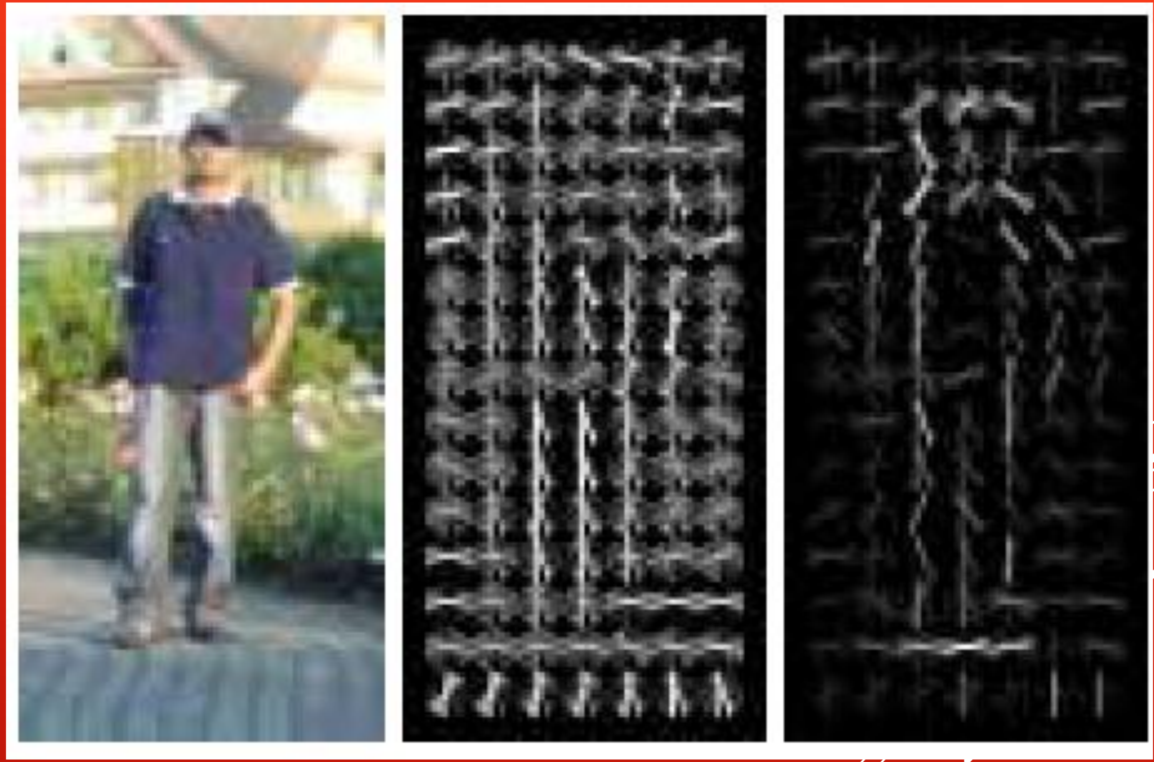
ECCV 12

Bharath Hariharan, Jitandra Malik, and Deva Ramanan

MOTIVATION

- State-of-the-art Object Detection
  - HOG
  - Linear SVM

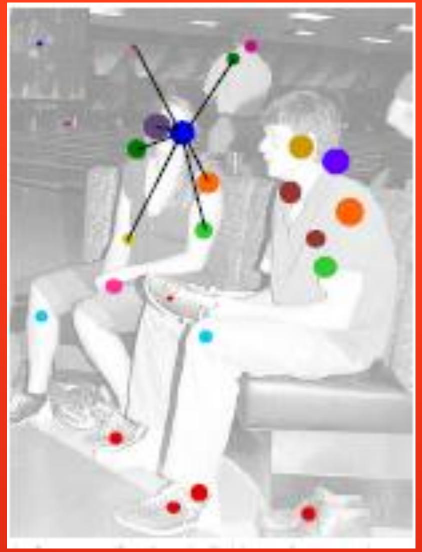




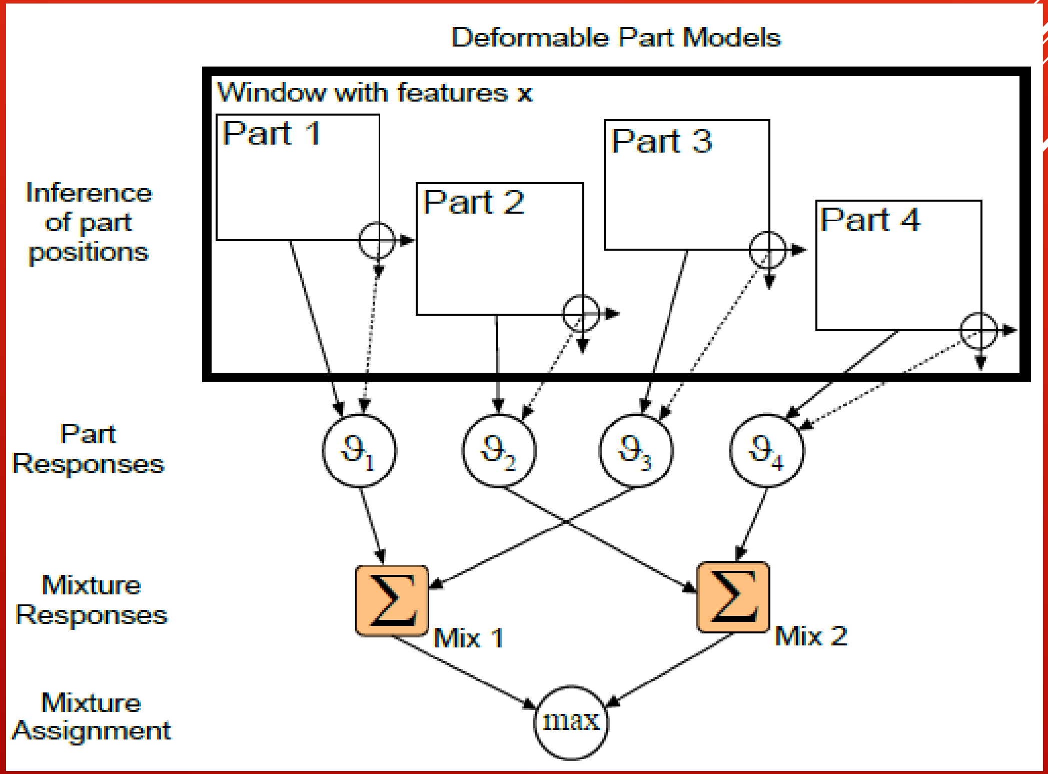
Dalal&Triggs Histograms of Oriented Gradients for Human Detection CVPR05

MOTIVATION

$$S_{\mathbf{w}}(x) = \max_{v \in \mathcal{V}, \Delta\theta} f(v, \Delta\theta)$$



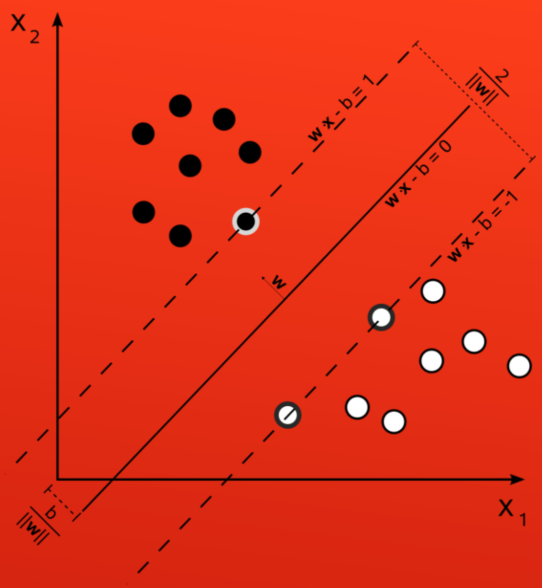
- State of the art Object Detection
  - Mixture models
  - Parts Based Models
- Train one linear for each class/mixture/part
  - Hundreds of objects of interest
  - Hard negative mining
  - Increasing number of mixtures (exemplar SVMs)
  - Increasing number of parts

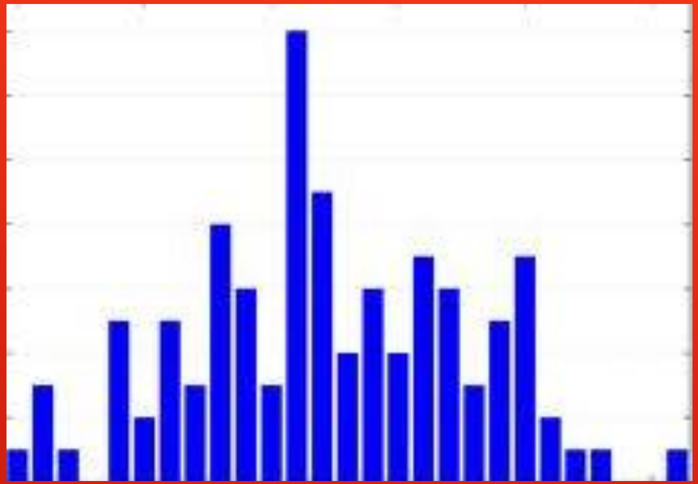


Can we make the training fast (closed form, no iteration) without losing much in performance?

P. Ott and M. Everingham. "Shared Parts for Deformable Part-based Models" CVPR11
Lubomir Bourdev, et al. ,Detecting People Using Mutually Consistent Poselet Activations ECCV10

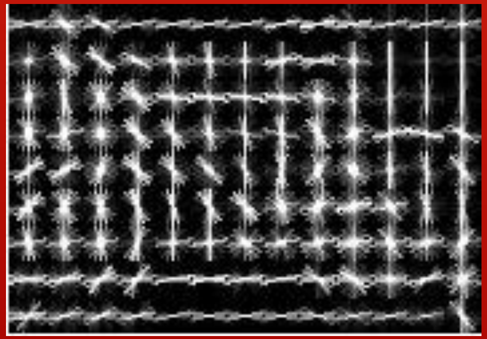
MOTIVATION

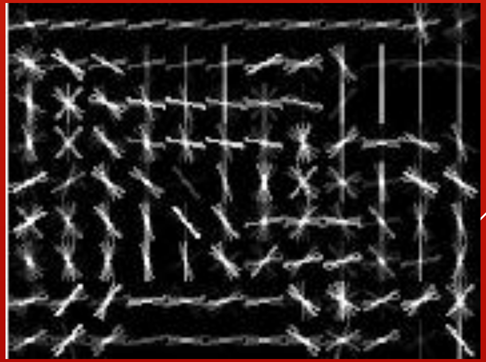




- What does linear SVMs do?
- Discriminative Feature selection!
- Can we do it closed form?
- Seems possible with LDA


Image


HOG


SVM


LDA

- Motivation
- LDA
- Closed form solution
- Parameters reduction/Regularization
- Properties of the model
- Clustering
- Results

CONTENTS

LINEAR DISCRIMINANT ANALYSIS (LDA) – BINARY CASE

- Training set (X,Y)
- Assumes normal $N(x; \mu_y, \Sigma)$ conditional distributions P(x | y=0) and P(x | y=1)
- Assumes equal class covariances
- From log of likelihood ratios give
  - $w = \Sigma^{-1}(\mu_1 - \mu_0)$
  - Dot product decision function: $\mathrm{f}(x) = w.x < c$

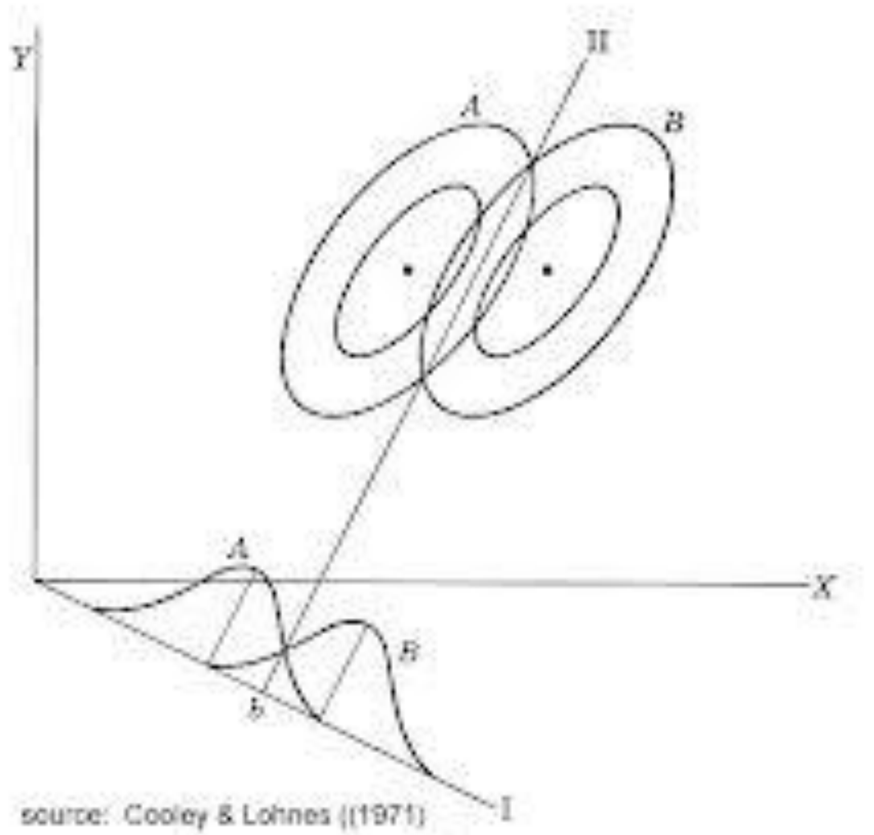


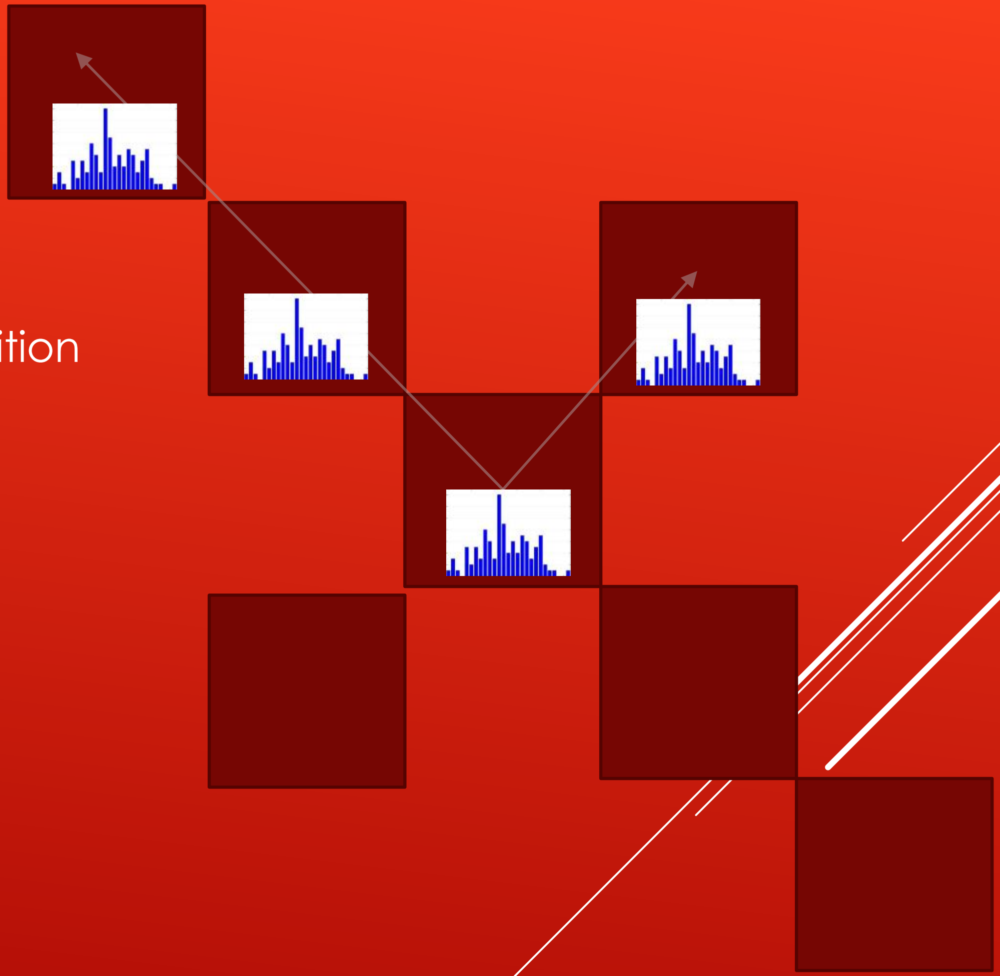
Figure 9.1.

CLOSED FORM SOLUTION

- Calculate ***object independent*** μ_c and Σ
  - Assume we want the same Σ across different classes
    - MLE of Σ becomes the covariance computed over all training samples
  - Lets assume number of negative samples are very much larger than positives of each class
    - μ_c can be computed once on all training samples including positives
- Even so, Number of parameters in a HOG template can be very large making estimation of a Σ very hard

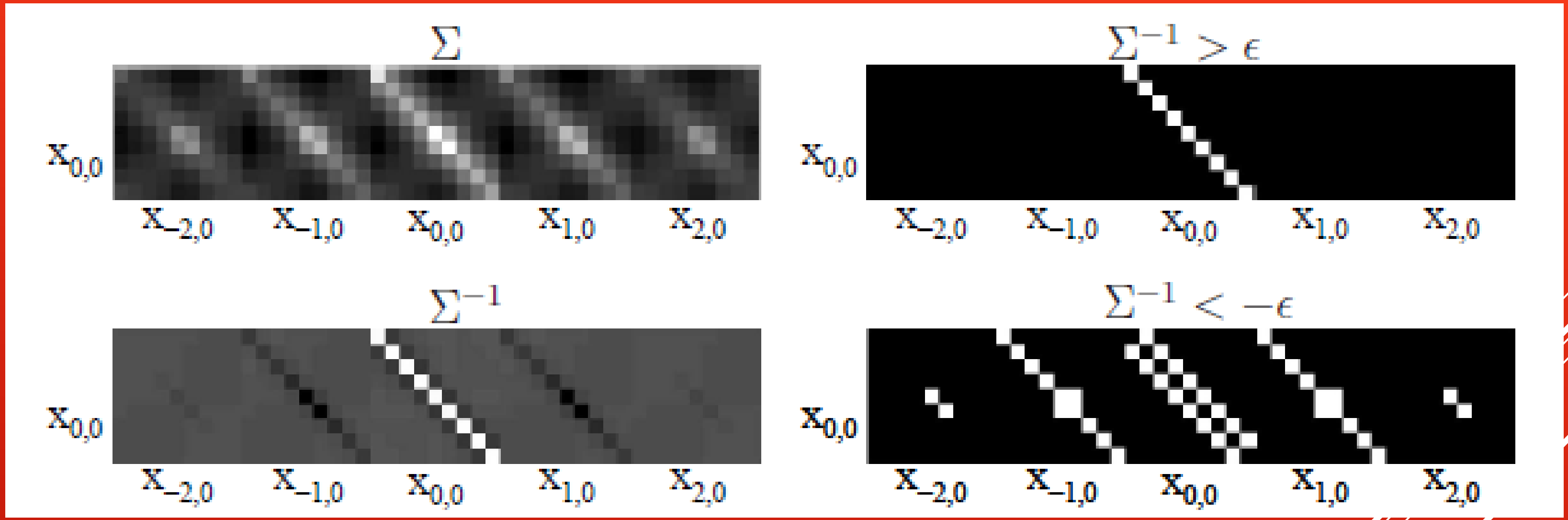
REGULARIZATION

- Translation and scale invariance
  - Take *all* the negative windows from each scale and position
- Stationary process
  - Mean will be computed for all HOG cells the same
  - Covariance will only model relative offsets
  - $N_0 d$ parameters only
- Covariance only on different cells offsets
- Still low rank and not invertible
  - Add a small amount to the diagonal (0.01)



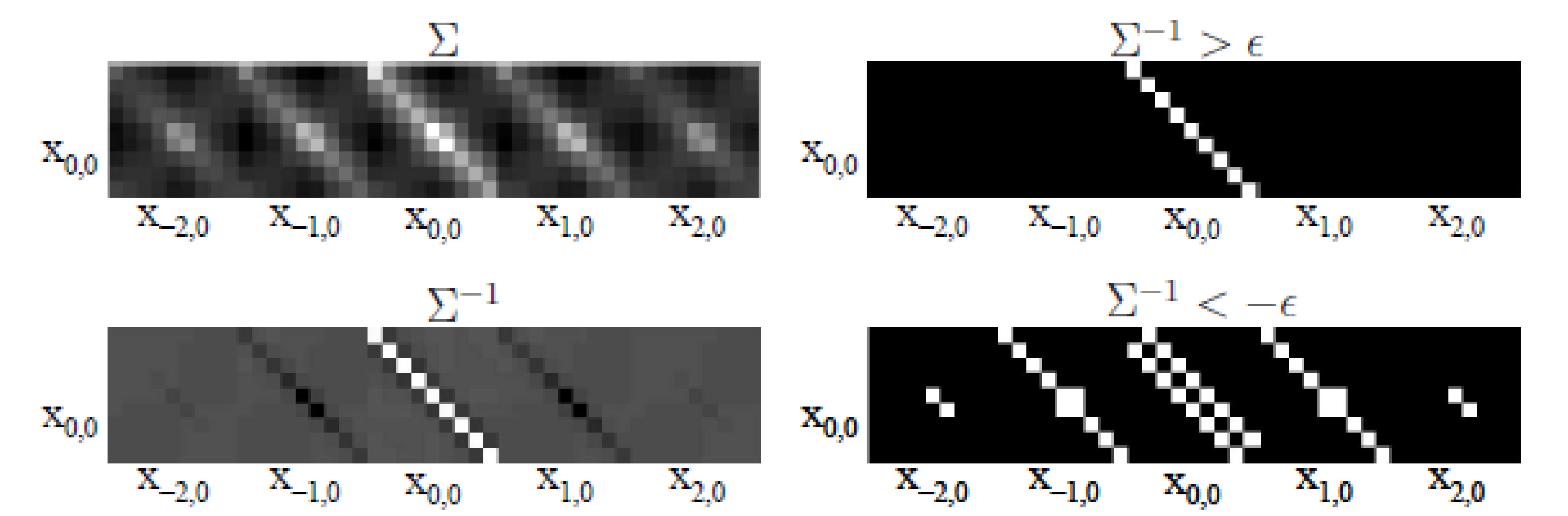
PROPERTIES OF THE COVARIANCE

- Structure of Σ
  - encodes spatial correlations between oriented gradients
- Sparsity



- Covariance of 9 orientation dimensions of one cell to 4 horizontal neighbors
- Dark is negative and light is positive
- Precision matrix is sparse
- Fades after a few cells away

PROPERTIES OF THE COVARIANCE



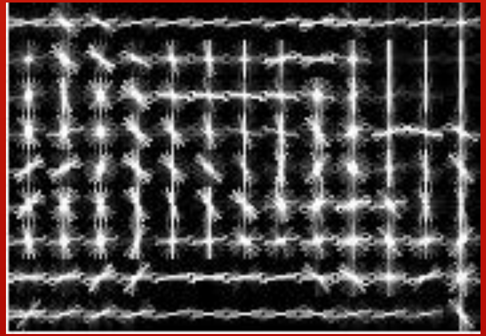
- Structure of Σ
  - encodes spatial correlations between oriented gradients
- Sparsity

- Covariance of 9 orientation dimensions of one cell to 4 horizontal neighbors
- Dark is negative and light is positive
- Precision matrix is sparse
- Fades after a few cells away

PROPERTIES OF THE COVARIANCE

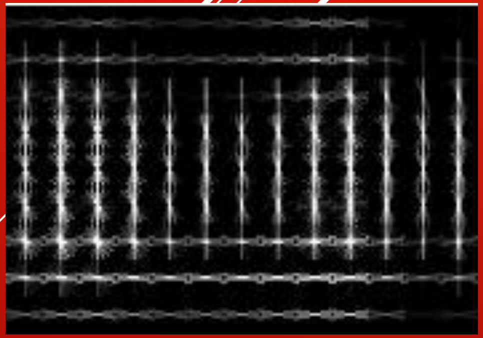
- Transformed (Whitened) HOG Feature (WHO)
  - $x' = \Sigma^{-\frac{1}{2}}(x - \mu_0)$
  - More meaningful distances
- Comparison to PCA
  - PCA keeps the dimensions with most variance among training samples
    - Removes the discriminative information!

Image



HOG

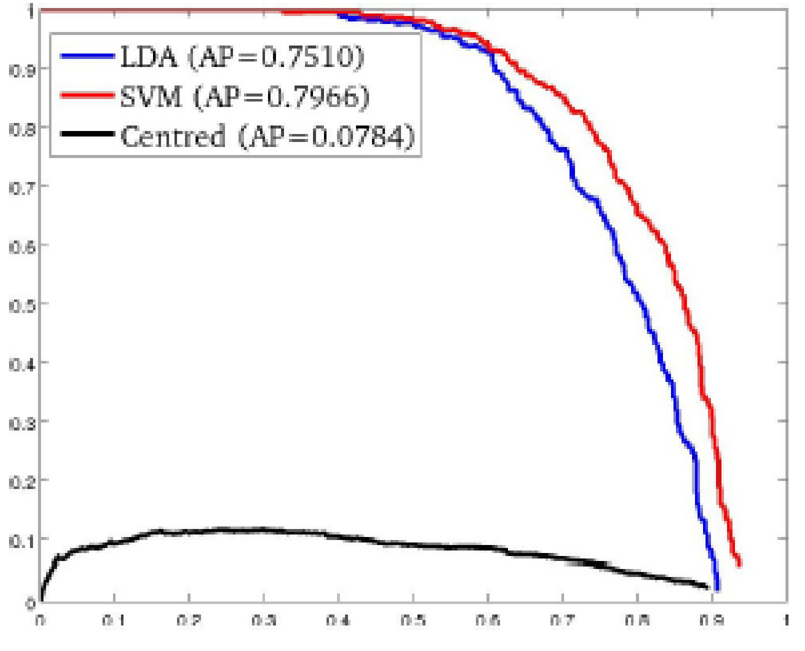
LDA



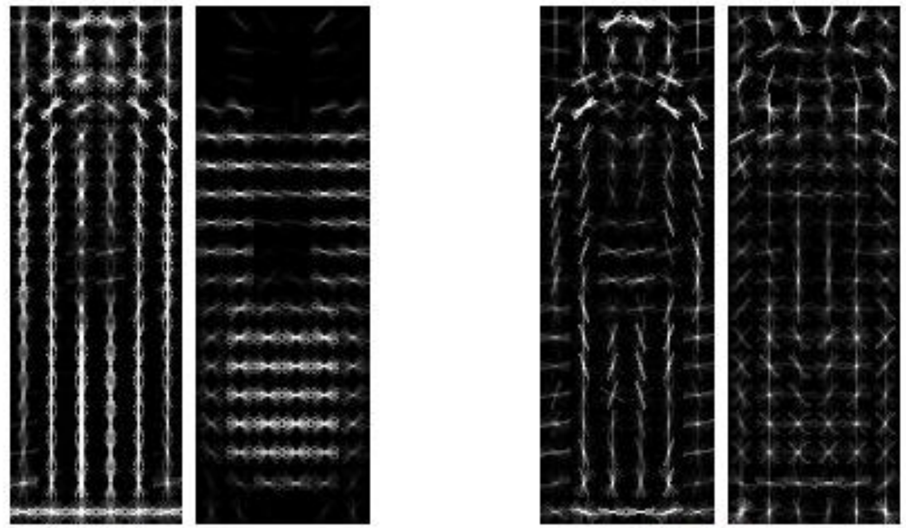
PCA

RESULTS – PEDESTRIAN DETECTION

- Comparing to Dalal Triggs HOG-SVM
  - With hard negative mining
- Method
  - Averaging: Compute positive average HOG (μ_1)
  - Centering: Subtract μ_0
  - Whitening: Multiply by Σ^{-1}
- To check the whitening effect use the centered version only
- Dalal&Triggs: 79.66%
- LDA: 75.10%
- Centered: 8%
- Pedestrians are well aligned



(a) AP


(b) Centered (c) LDA

CLUSTERING

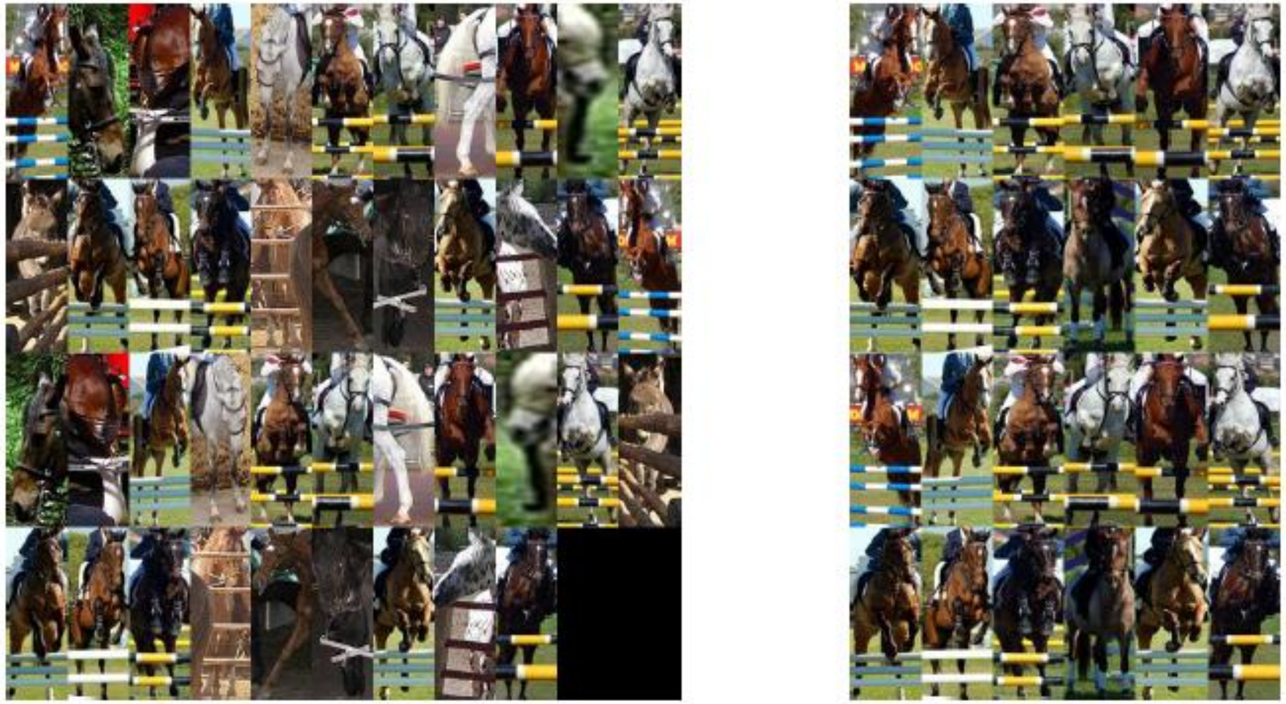
- Use transformed feature
- Recursive normalized cuts

(a) HOG

(b) PCA

(c) WHO

WHO CLUSTERING RESULTS



(a) horse

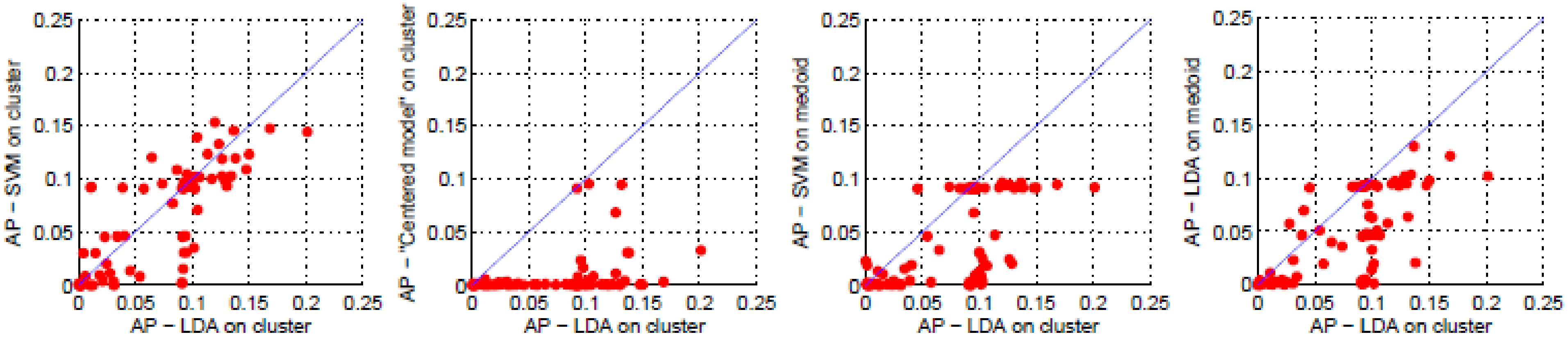
(b) aeroplane

SOME MINOR RESULTS

- Average over all clusters performance
- Airplane, bus, horse
- Correlation of SVM and LDA
- Low training computation
- Still slow in testing phase

	LDA on cluster	SVM on cluster	LDA on medoid	SVM on medoid	Centered
Mean AP	7.59 ± 4.86	6.75 ± 4.80	4.84 ± 4.13	4.05 ± 4.12	0.74 ± 2.02
Median AP	9.25 ± 3.86	9.16 ± 4.04	4.65 ± 3.71	2 ± 3.6	0.06 ± 0.7

Table 1. Mean and median AP (in %) of the different models.



COMBINING ACROSS CLUSTERS

- 1- Use Exemplar SVMs method
- 2- change the SVM to an LDA with object independent background model
- 3- To make it fast in test
  - Use clustering train a classifier for each cluster
    - Train an LDA for each cluster
    - Produce Feature Vector1: Dot product of a window WHO feature to all exemplars WHO
    - Produce Feature Vector 2: Dot product of a window WHO to all clusters WHO
    - Train a linear SVM on the concatenated feature

	ESVM +Calibr	ESVM +Co-occ	ELDA +Calibr	Ours-only 1	Ours-only 2	Ours-full
aeroplane	20.4	20.8	18.4	17.4	22.1	23.3
bicycle	40.7	48.0	39.9	35.5	37.4	41.0
bird	9.3	7.7	9.6	9.7	9.8	9.9
boat	10.0	14.3	10.0	10.9	11.1	11.0
bottle	10.3	13.1	11.3	15.4	14.0	17.0
bus	31.0	39.7	39.6	17.2	18.0	37.8
car	40.1	41.1	42.1	40.3	36.8	38.4
cat	9.6	5.2	10.7	10.6	6.5	11.5
chair	10.4	11.6	6.1	10.3	11.2	11.8
cow	14.7	18.6	12.1	14.3	13.5	14.5
diningtable	2.3	11.1	3	4.1	12.1	12.2
dog	9.7	3.1	10.6	1.8	10.5	10.2
horse	38.4	44.7	38.1	39.7	43.1	44.8
motorbike	32.0	39.4	30.7	26.0	25.8	27.9
person	19.2	16.9	18.2	23.1	21.3	22.4
pottedplant	9.6	11.2	1.4	4.9	5.1	3.1
sheep	16.7	22.6	12.2	14.1	13.8	16.3
sofa	11.0	17.0	11.1	8.7	12.2	8.9
train	29.1	36.9	27.6	22.1	30.6	30.3
tvmonitor	31.5	30.0	30.2	15.2	12.8	28.8
Mean	19.8	22.6	19.1	17.0	18.3	21.0

WHAT WE WANTED! TO DO

- Train exemplar SVMs for each sample
- Take the selected feature as the new transformed feature
  - Do Image Classification (combined features and nonlinear SVMs)
  - Clustering and Linear SVMs ALMOST DONE
  - New similarity measure for verification (like face verification in LFW)
  - Model Natural images with some distribution to make training Exemplar SVMs fast DONE
    - One can use more sophisticated modelling of natural patches (GMM))
  - Make a pyramid of second, 3rd, … order discriminant features