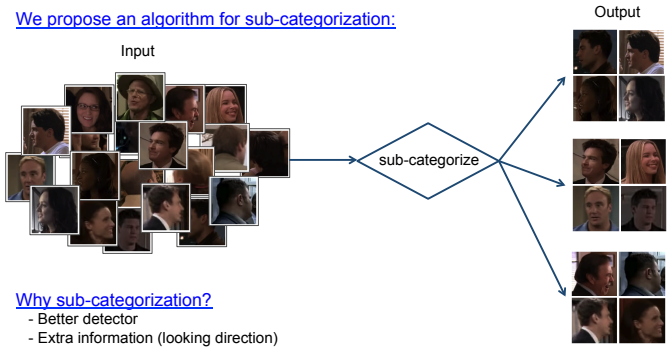
Discriminative Sub-categorization

Minh Hoai, Andrew Zisserman
Visual Geometry Group, University of Oxford

Objective

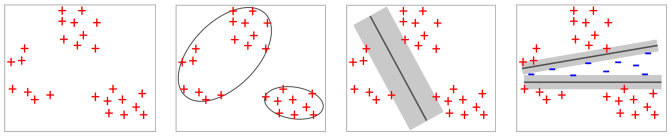
We propose an algorithm for sub-categorization:



Why sub-categorization?
- Better detector
- Extra information (looking direction)

Clustering Approaches

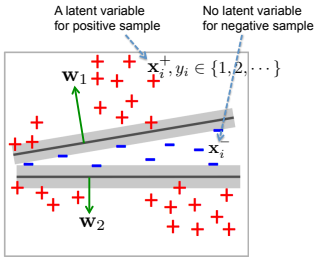
Clustering approaches:



No sub-categorization | *k*-means | Max-margin clustering (e.g., Xu *et al.* '04, Hoai & De la Torre '12) | MI-SVM, LSVM (e.g., Andrews *et al.* '03, Felzenszwalb *et al.* '10)

Suitable for object detection (require separation between positive and negative data)

Review of MI-SVM, LSVM



Objective:
- Optimize SVM parameters $\{\mathbf{w}_j\}$
- Determine latent variables $\{y_i\}$

Iterative optimization, alternating:
- Given $\{\mathbf{x}_i^+ | y_i = j\}$ and $\{\mathbf{x}_i^-\}$

  Solve SVM $\{\mathbf{w}_j\}$
- Fix $\{\mathbf{w}_j\}$, update latent variables

  $y_i = \underset{j}{\operatorname{argmax}}\, \mathbf{w}_j^T \mathbf{x}_i^+$

Often leads to cluster degeneration: a few clusters claim most data points

An explanation (not rigorous proof) for cluster degeneration:

Suppose Cluster 1 has many more members than Cluster 2
⇒ It is much harder to separate Cluster 1 from negative data
⇒ Cluster 1 has a much smaller margin

$$\Rightarrow \frac{1}{||\mathbf{w}_1||} \ll \frac{1}{||\mathbf{w}_2||} \Rightarrow ||\mathbf{w}_1|| \gg ||\mathbf{w}_2|| \Rightarrow \mathbf{w}_1^T \mathbf{x}^+ > \mathbf{w}_2^T \mathbf{x}^+$$

⇒ Cluster 1 will get bigger
⇒ The big gets bigger

Discriminative Sub-categorization

Objective functions

Change from the MI-SVM/LSVM formulation:

$$\underset{\{\mathbf{w}_j\},\{y_i\}}{\text{minimize}} \sum_{j=1}^{k} \frac{1}{k} ||\mathbf{w}_j||^2 + \boxed{\text{Margin violation}}$$

To this formulation (called DSC)

$$\underset{\{\mathbf{w}_j\},\{y_i\}}{\text{minimize}} \sum_{i=1}^{n} \frac{1}{n} ||\mathbf{w}_{y_i}||^2 + \boxed{\text{Margin violation}}$$

Coupled with latent variable

This is equivalent to

$$\underset{\{\mathbf{w}_j\},\{y_i\}}{\text{minimize}} \sum_{j=1}^{k} \frac{n_j}{n} ||\mathbf{w}_j||^2 + \boxed{\text{Margin violation}}$$

n_j: Proportion of samples in Cluster j

k: # of clusters
n: # of positive samples
y_i : cluster assignment
$\mathbf{w}_j$: SVM parameter

Cluster Assignment

Change from MI-SVM/LSVM formulation:

$$y_i = \underset{j}{\operatorname{argmax}}\, \mathbf{w}_j^T \mathbf{x}_i^+$$

To DSC formulation:

$$y_i = \underset{j}{\operatorname{argmin}} \left\{ ||\mathbf{w}_j||^2 + C \max(0, 1 - \mathbf{w}_j^T \mathbf{x}_i^+) \right\}$$

Similarity between DSC and K-means:

$$y_i = \underset{j}{\operatorname{argmin}} ||\mathbf{x}_i^+ - \mathbf{w}_j||^2$$

$$y_i = \underset{j}{\operatorname{argmin}} \left\{ ||\mathbf{w}_j||^2 - 2\mathbf{w}_j^T \mathbf{x}_i^+ \right\}$$

Optimization algorithm

- Uses block-coordinate descent: alternate between updating cluster assignment and updating SVM parameters.
- Updating SVM parameters requires quadratic programming. We uses stochastic gradient descent in our implementation.

Clustering performance

- Datasets: UCI datasets + MNIST
- Performance measure: cluster purity, agreement between clusters with class labels

Dataset	#classes	#features	#points	k-means	LSVM	DSC (ours)
Gas Sensor	6	128	13910	46.38 ± 0.69	56.74 ± 1.88	**60.82 ± 1.64**
Landsat	6	36	4435	**78.72 ± 2.08**	69.37 ± 2.32	**76.73 ± 2.38**
Segmentation	7	19	2310	71.96 ± 1.75	65.89 ± 2.36	**74.41 ± 1.85**
Steel Plates	7	27	1941	**53.29 ± 1.51**	**52.64 ± 2.02**	**54.60 ± 1.98**
Wine quality	7	12	4898	43.43 ± 1.58	**55.00 ± 2.35**	**54.21 ± 1.65**
Digits	10	64	5620	76.38 ± 1.72	77.83 ± 1.57	**80.15 ± 1.18**
Semeion	10	256	1593	64.64 ± 1.20	64.32 ± 1.58	**66.74 ± 1.43**
MNIST	10	784	60000	**65.38 ± 1.43**	63.99 ± 1.36	**66.18 ± 1.34**
Letter	26	16	20000	33.35 ± 0.48	40.27 ± 0.88	**44.38 ± 0.74**
Isolet	26	617	6238	62.15 ± 1.22	61.95 ± 1.22	**64.08 ± 1.18**
Amazon Reviews	50	10000	1500	**24.93 ± 0.32**	**24.89 ± 0.41**	**25.08 ± 0.38**

Results within one standard error of the maximum value are printed in bold

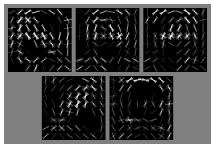
Head Sub-categorization

Qualitative evaluation

Input:
- 4040 head images from TVHI dataset
- 4872 negative patches from INRIA Person dataset

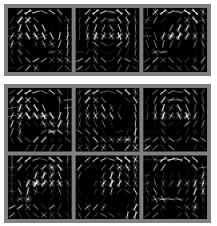
Output:

5 sub-categories (HOG weights)

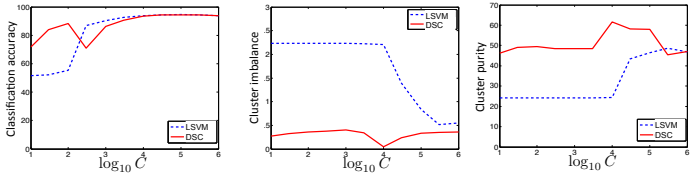


High score images

Low-score images

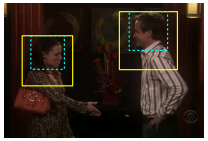


Quantitative Analysis



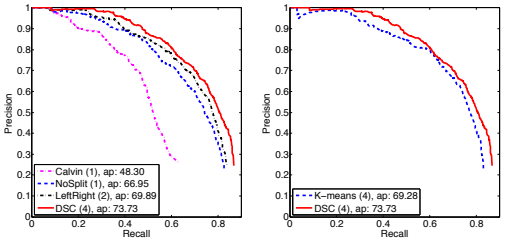
Upper-body Detection

- Uses DSC to initialize training DPM detector for upper bodies
- Dataset: positive examples from TVHI dataset, negative examples from INRIA Person dataset



Examples of upper bodies and heads

Precision-recall for upper body detection



Calvin: DPM with 1 component for near-frontal upper bodies
NoSplit: DPM with 1 component, i.e., no sub-categorization
LeftRight: DPM with 2 components, left and right facing
DSC: DPM with 4 components, initialized with DSC
K-means: DPM with 4 components, initialized with *k*-means on HOGs