Latent Hough Transform
for Object Detection

Nima Razavi Juergen Gall Pushmeet Kohli Luc Van Gool

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

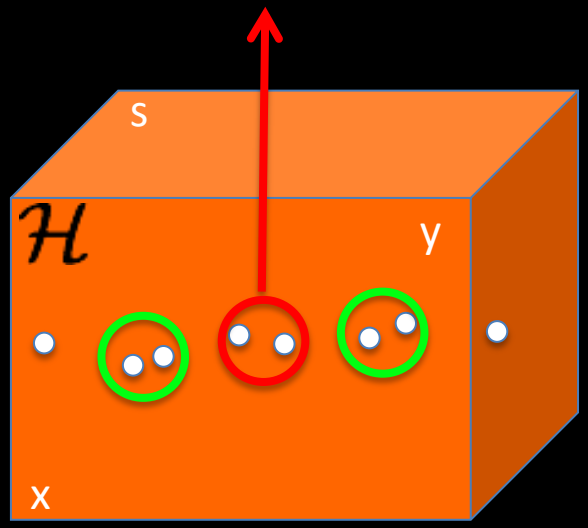
Microsoft Research

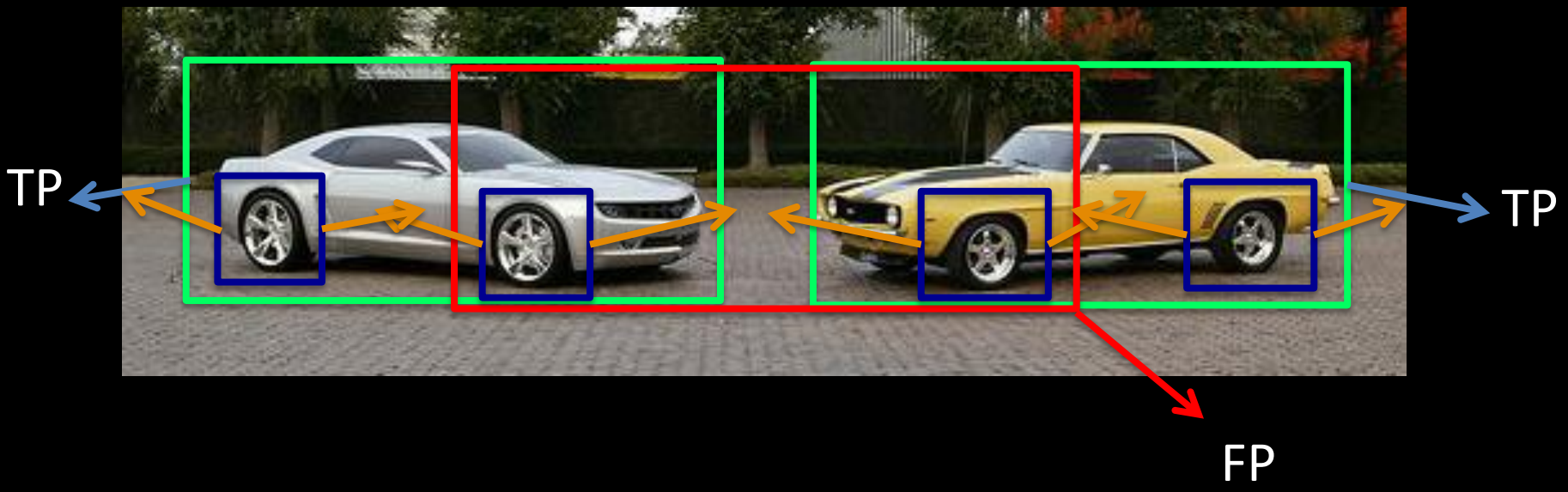
KATHOLIEKE UNIVERSITEIT
LEUVEN

Detection with the Hough Transform

Accumulation of inconsistent votes

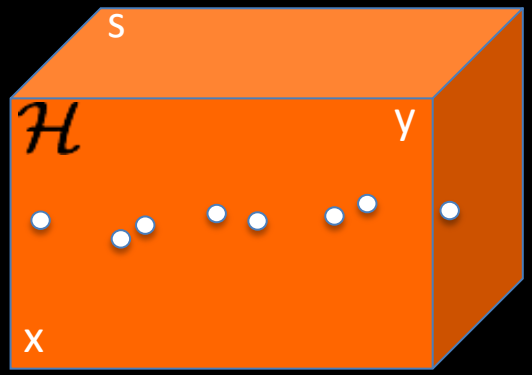
Hough Space
(position and scale)



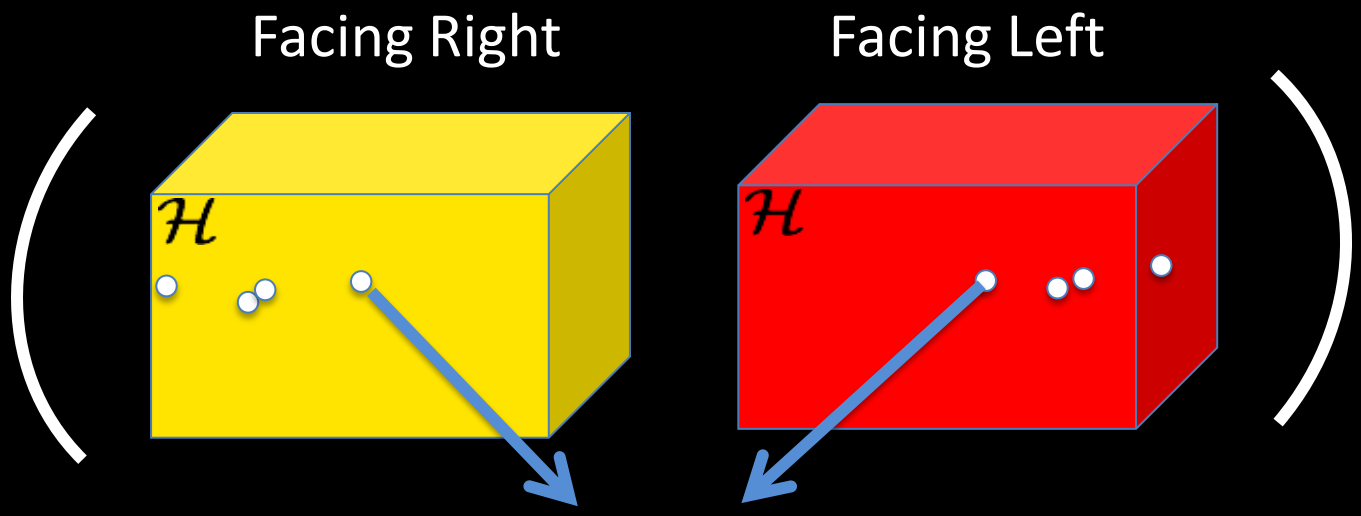


How to enforce consistency of votes?

- Voting for viewpoint?



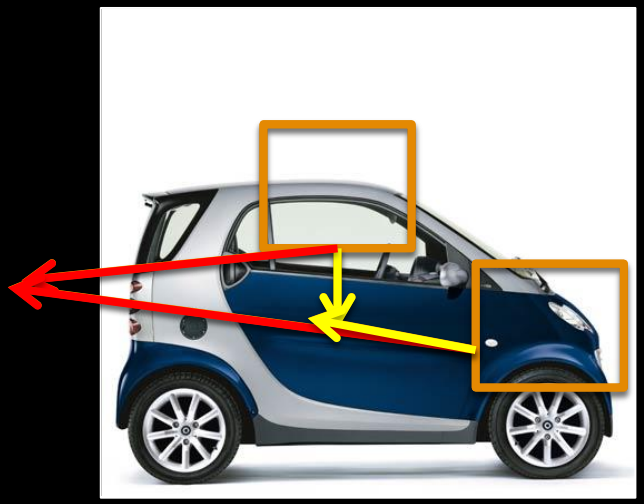
- Voting for viewpoint?



votes inconsistent in viewpoint are not accumulated

- Voting for type?

Smart
Limousine



What about color, aspect ratio, etc.?

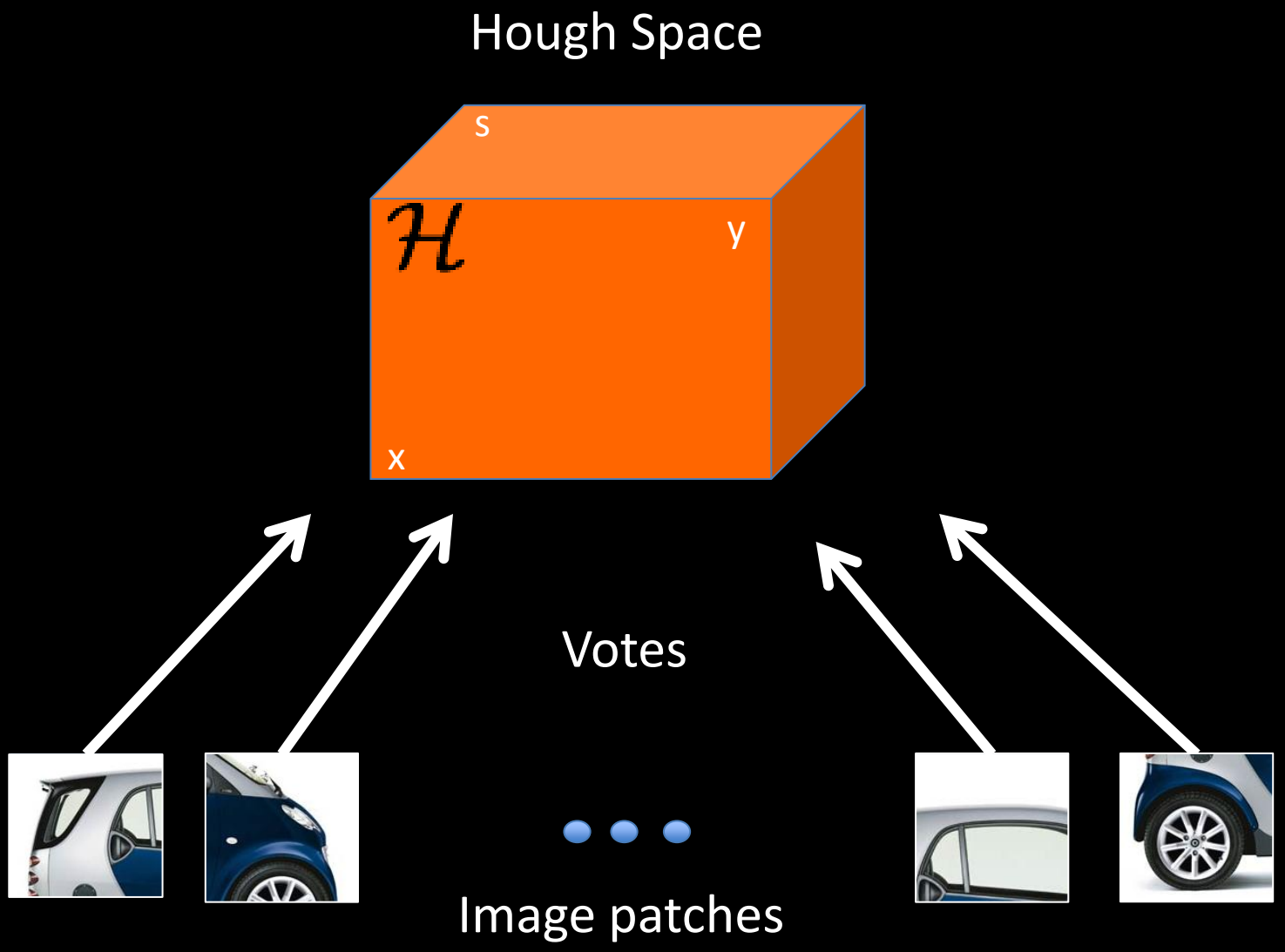
Previous Works

- Voting for other attributes:
  - pose (Seemann'07)
  - viewpoint (Thomas'06,Razavi'10)
  - depth (Sun'10)
  - shapes (Marszalek'08)
  - etc.

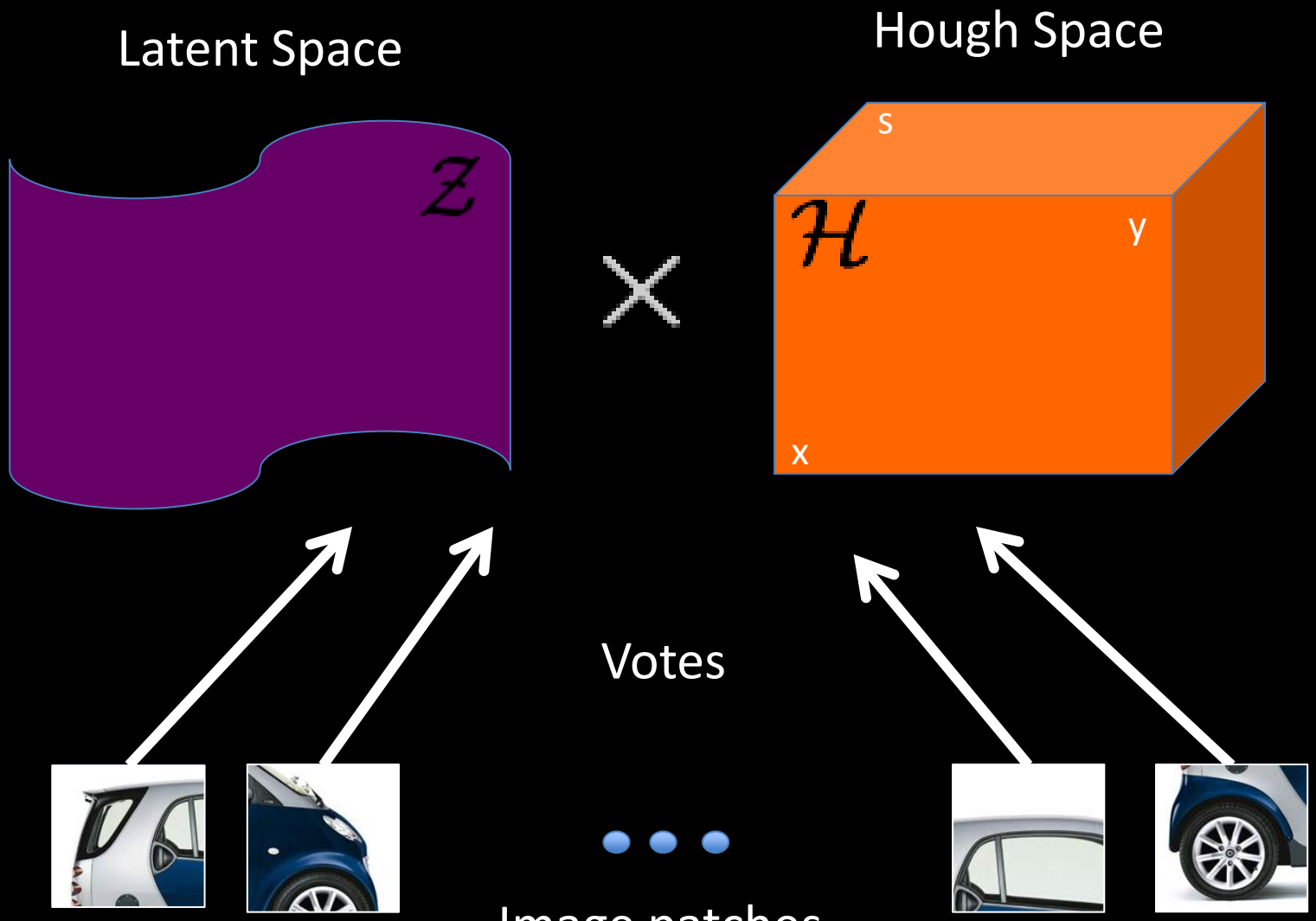
- But ….
  - What attribute to choose?
  - How to quantize it?
  - There is also a cost of annotations
  - We cannot use all attributes together
    - HT does not work well on high dimensions (Stephens'91)

Can we learn the attributes to be consistent over?

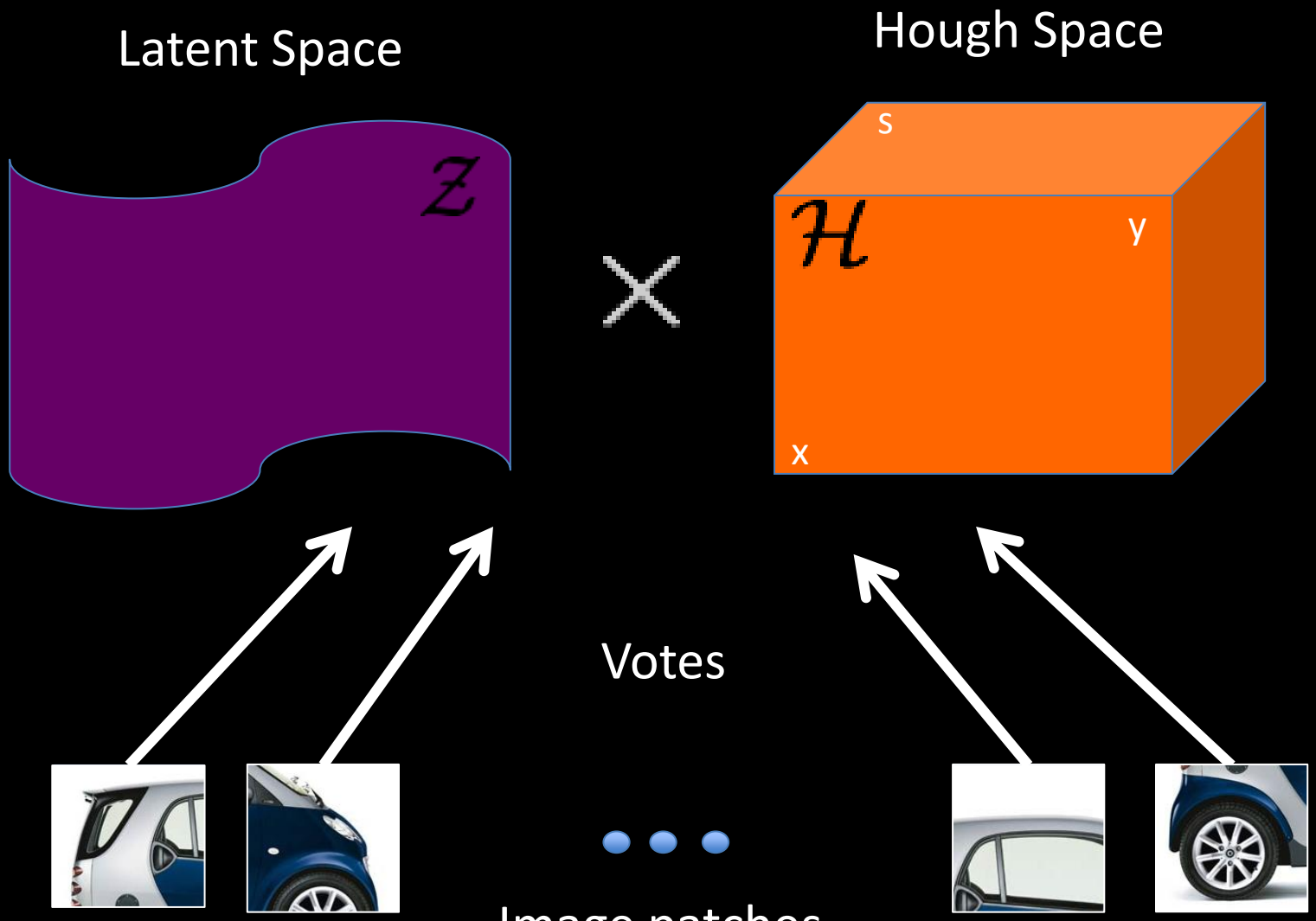
Hough Transform



Hough Transform



Latent Hough Transform



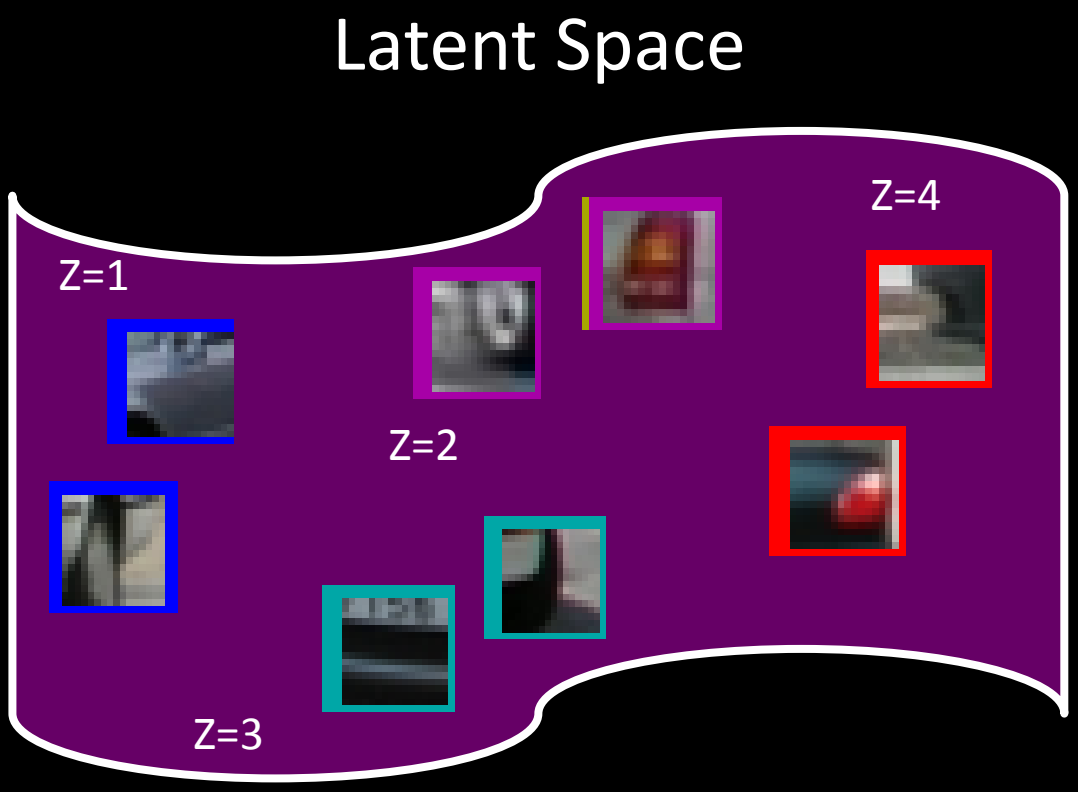
Latent Matrix

- Every vote is a patch in a training image (Leibe'08)

Training Image

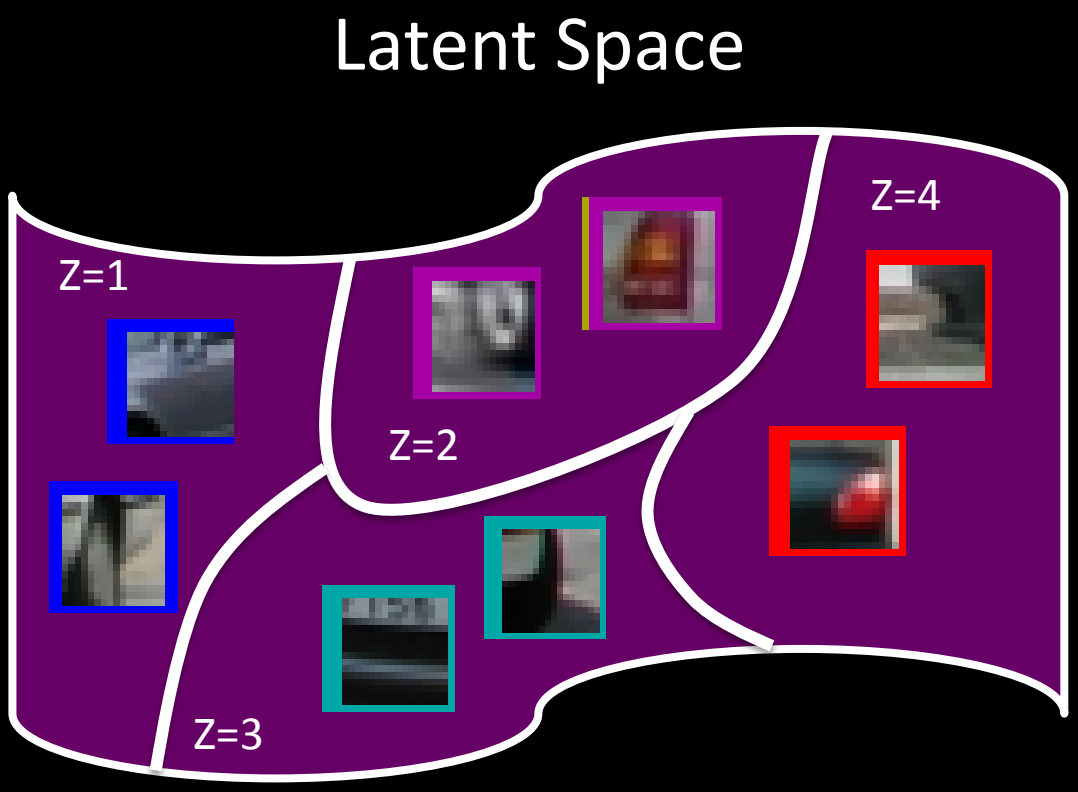
Latent Matrix

- Every vote is a patch in a training image (Leibe'08)



Latent Matrix

- Every vote is a patch in a training image (Leibe'08)



Latent Matrix

- Every vote is a patch in a training image (Leibe'08)
- A latent grouping can be represented as a matrix

Latent Matrix

- Every vote is a patch in a training image (Leibe'08)
- A latent grouping can be represented as a matrix

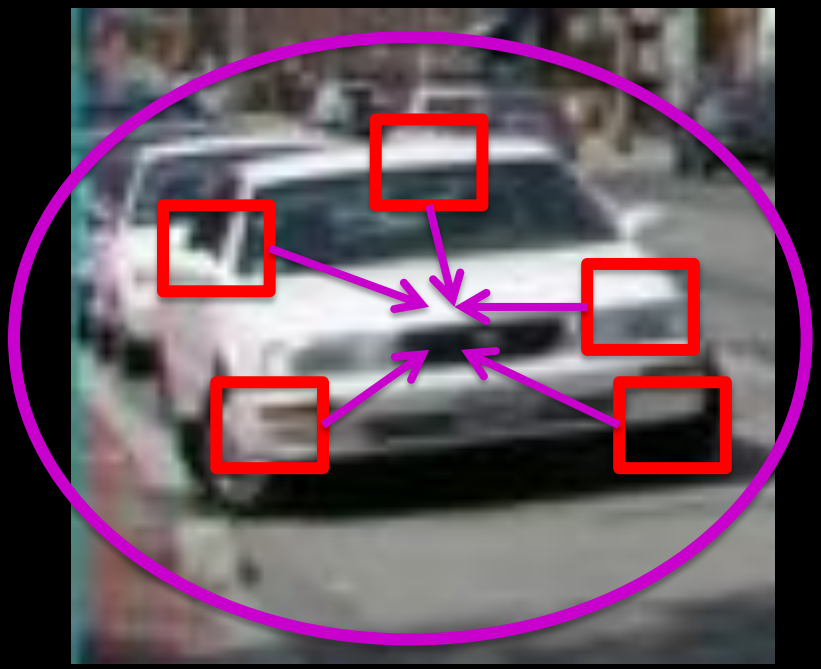
Latent Matrix

- Every vote is a patch in a training image (Leibe'08)
- A latent grouping can be represented as a matrix

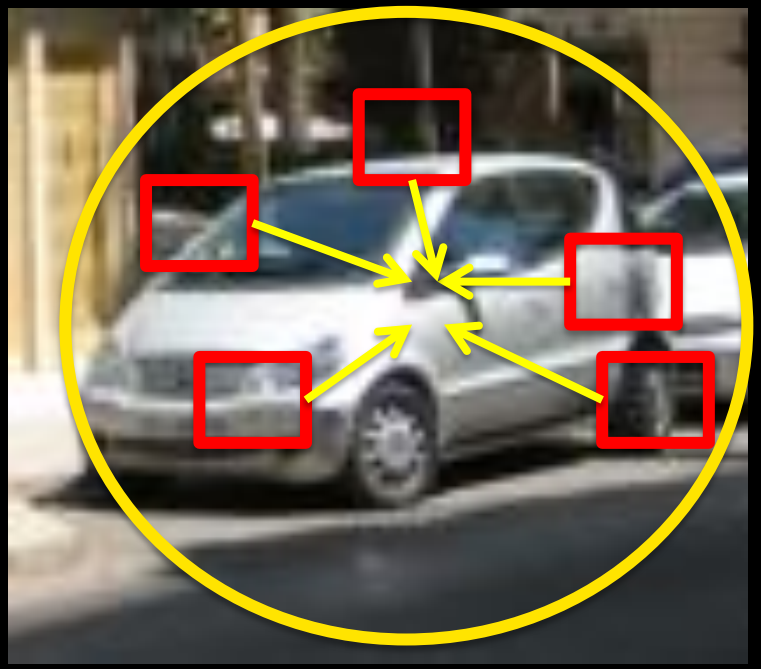
- The number of votes is very large (~1 M)

- The number of votes is very large (~1 M)
- Votes from the same training image are all consistent

  → we can pre-group them together (~1000)



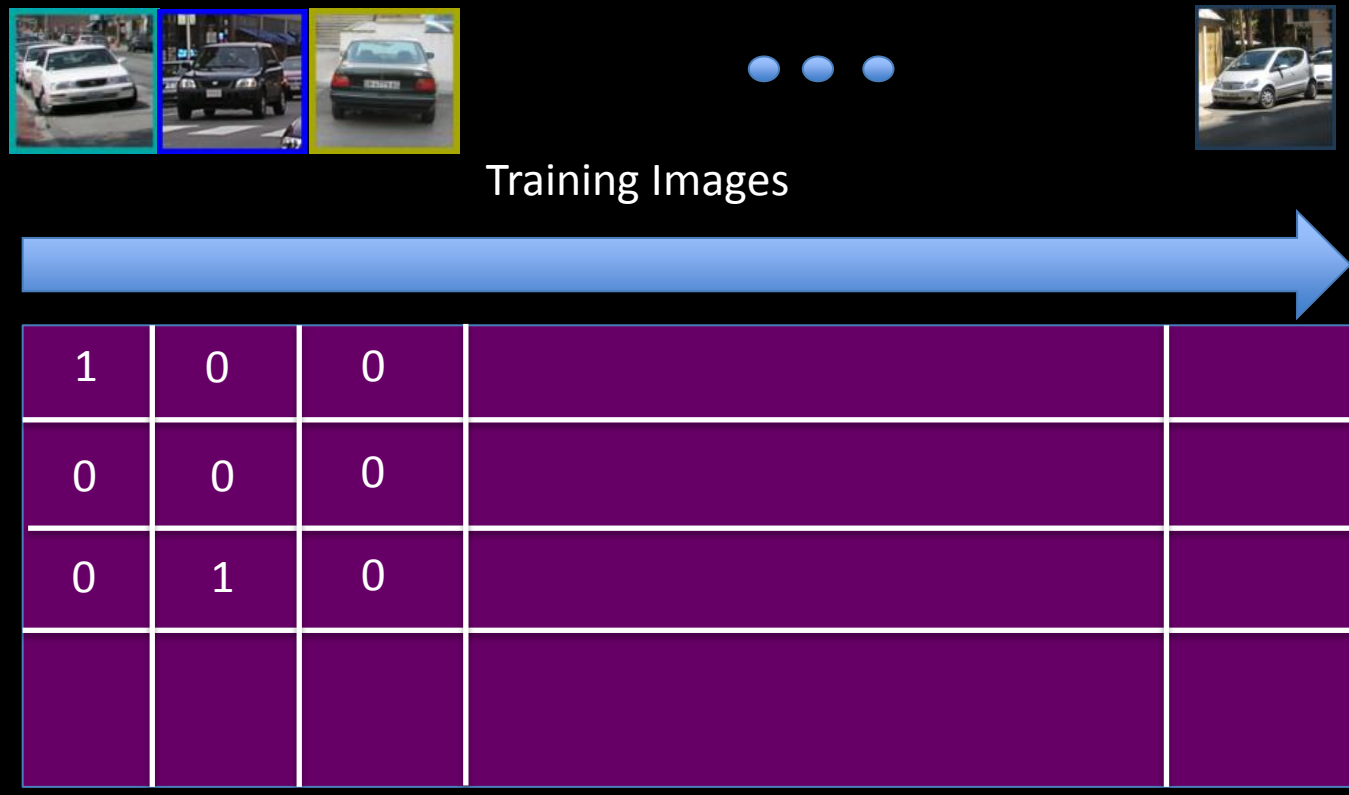
Training Image 1



Training Image 2

- The number of votes is very large (~1 M)
- Votes from the same training image are all consistent

  → we can pre-group them together (~1000)

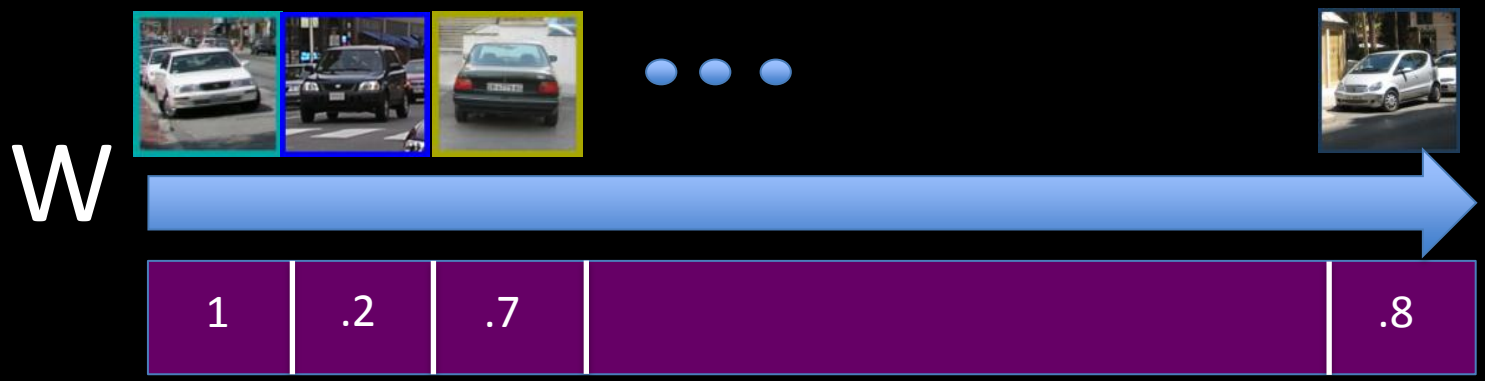


Interesting Special Cases of our Model

Special Cases of W

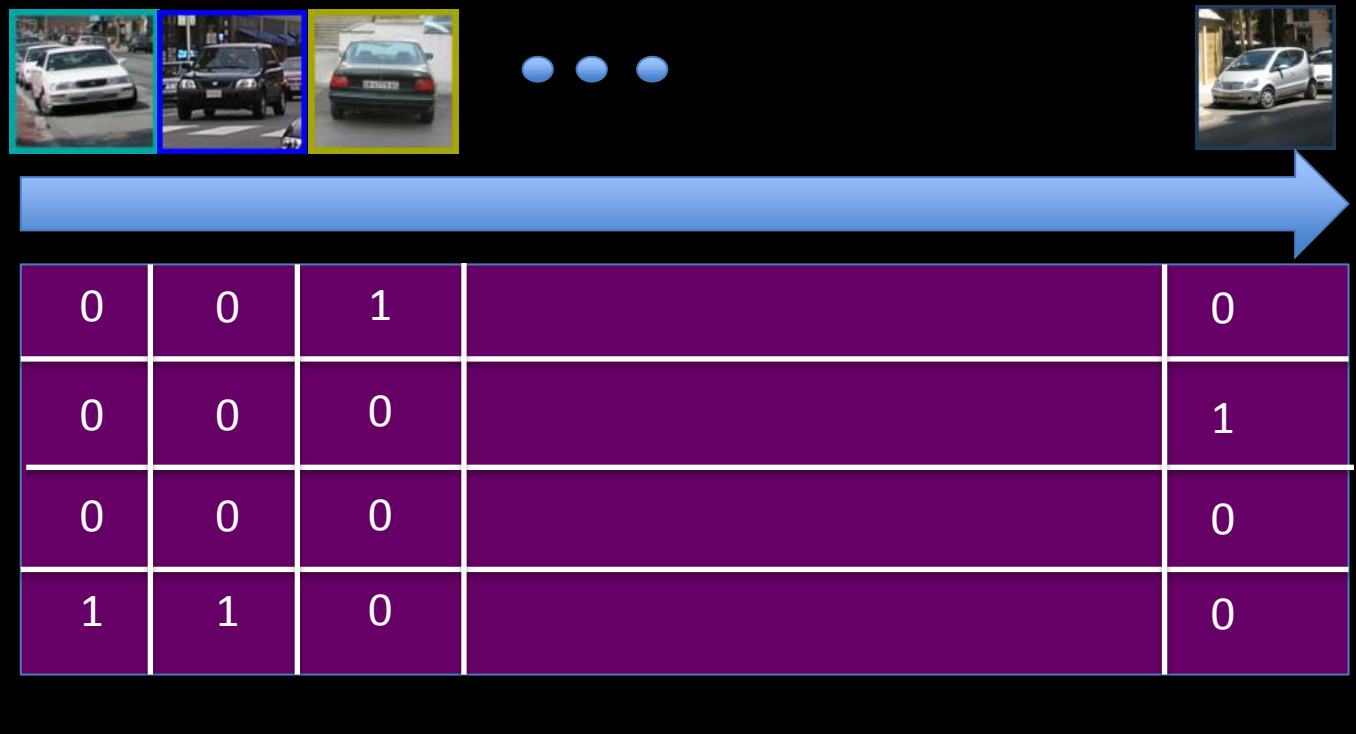
- Single Row
  - Hough transform with weighted training examples

    Related to Max Margin HT (Maji'09, Zhang'10)



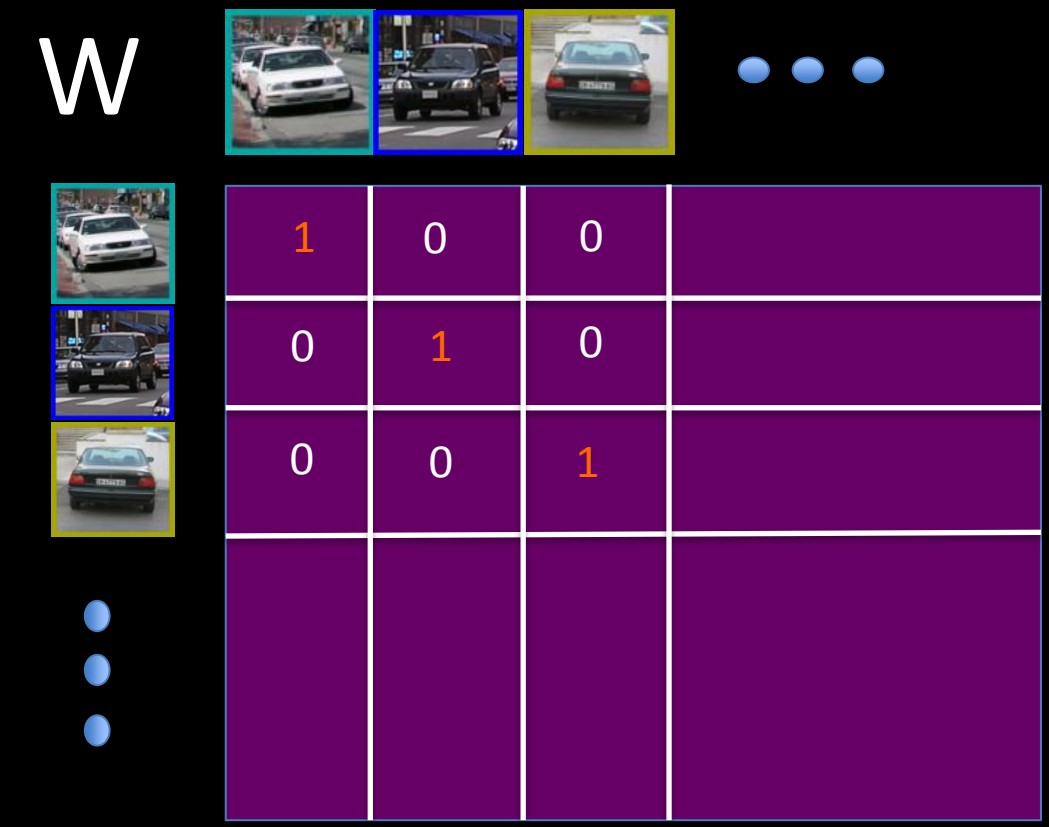
Special Cases of W

- Clustering/Annotations
  - Disjoint grouping with a {0,1} matrix
  - Related to Latent SVMs: (Felzenszwalb et al.'10)



Special Cases of W

- One training image per group:
  - Related to Exemplar-SVMs (Malisiewicz et al.'11)



Special Cases of W

- Uniform weights
  - Equals a single group

Discriminative Learning of W

$$\hat{W} = \arg\max_{W} O(W, R).$$

$O(W, R)$ Objective: average precision on the validation set R

Our objective is non-convex and not even continuous

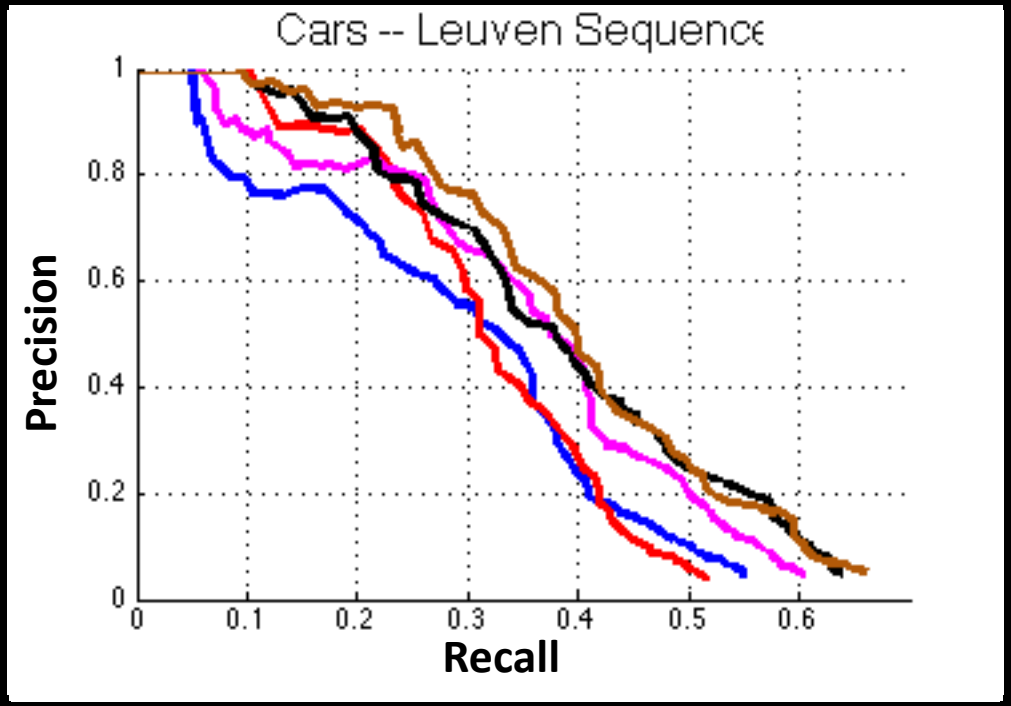
We do global optimization with a variation of simulated annealing

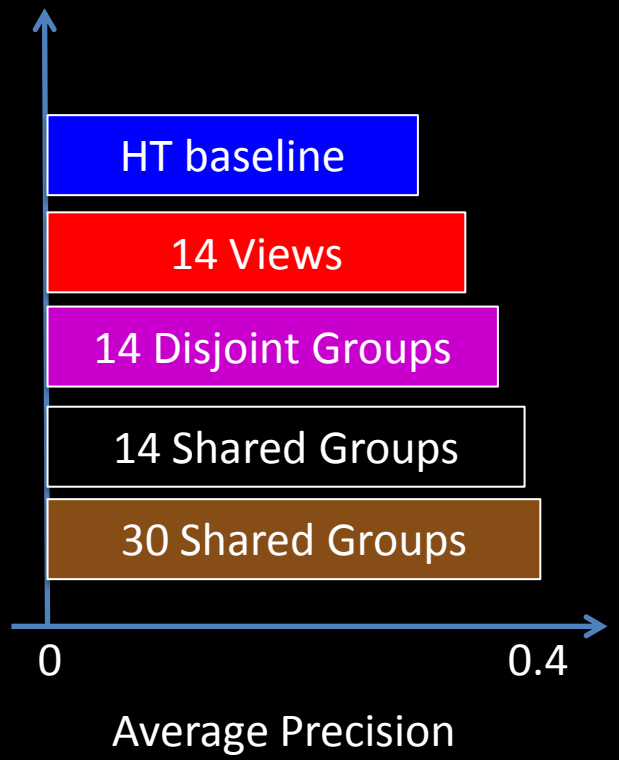
Experiments

- Setup
  - Two datasets:
    - ETHZ cars dataset (Leibe et al.'06)
    - PASCAL VOC 2007 (Everingham et al.'07)
  - Pre-train a codebook per category only once
    - Using Hough Forests (Gall and Lempitsky'09)
    - The codebook and the offset stay identical
  - Learning W using the validation set

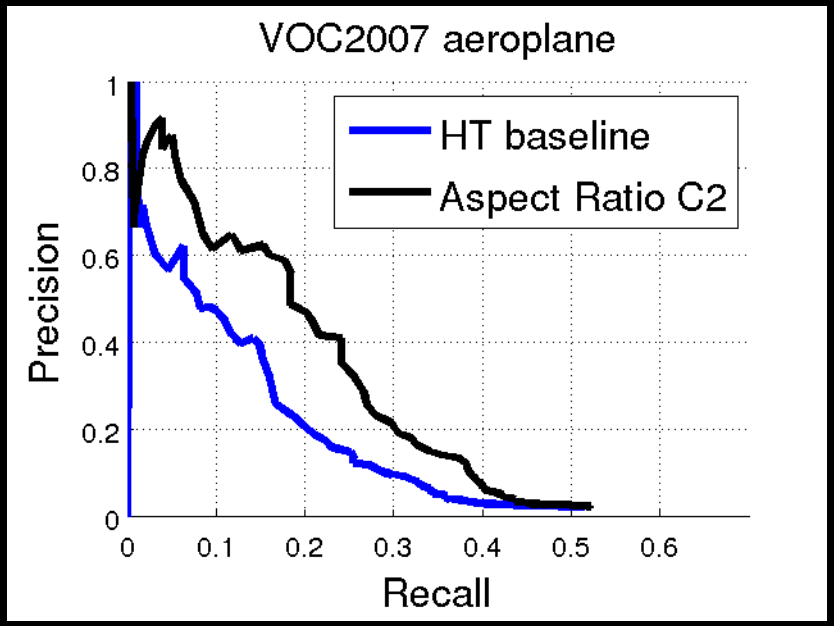
Learning or Annotation?

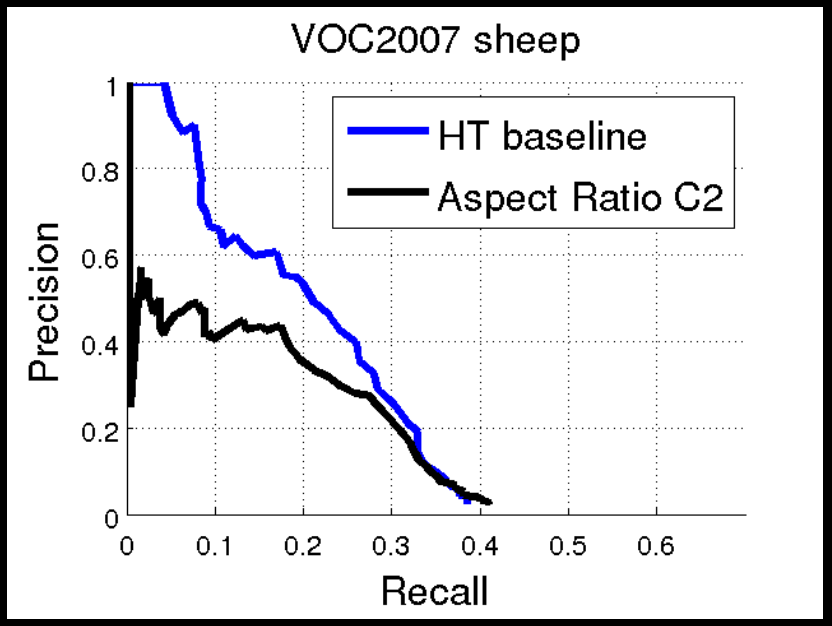
- ~3000 training images, annotated for 14 views
- Testing on Leuven video sequence (Leibe'07)



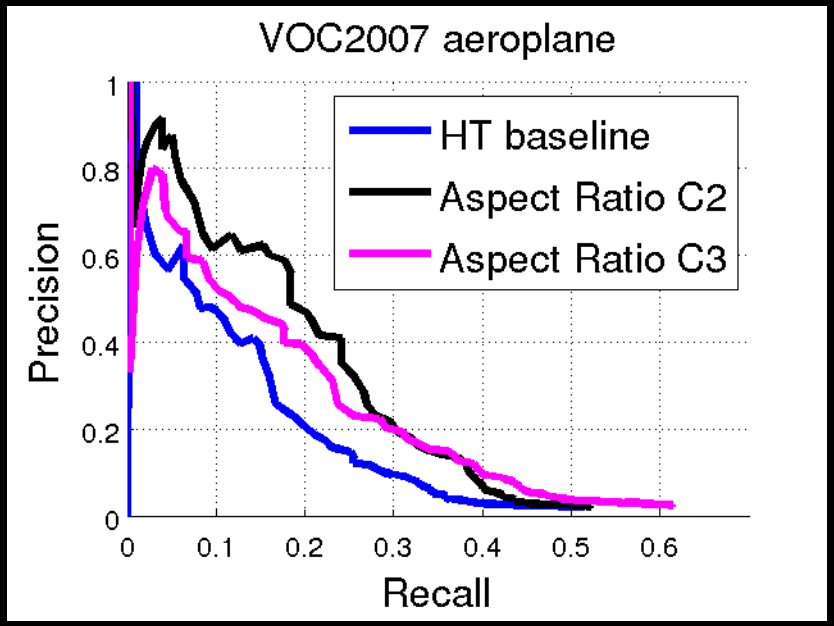


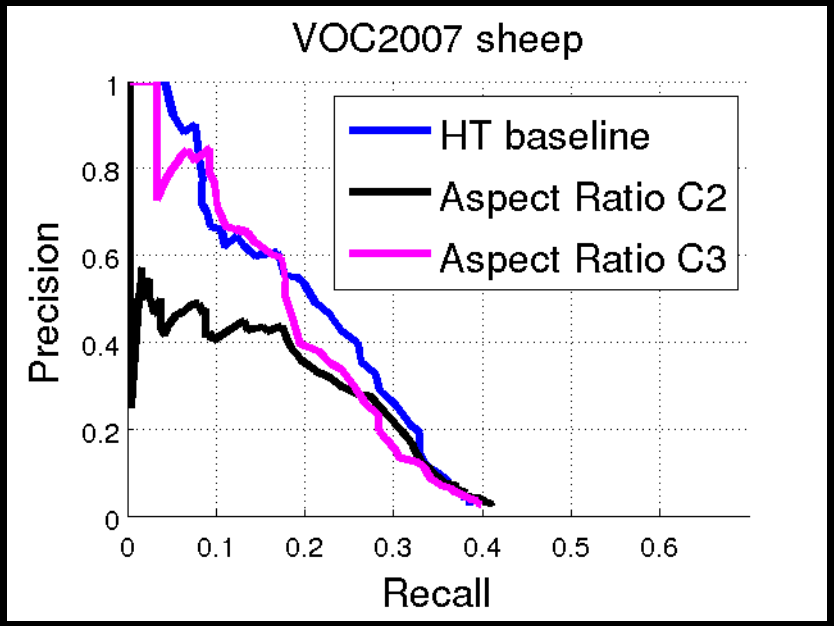
Learning or Clustering?



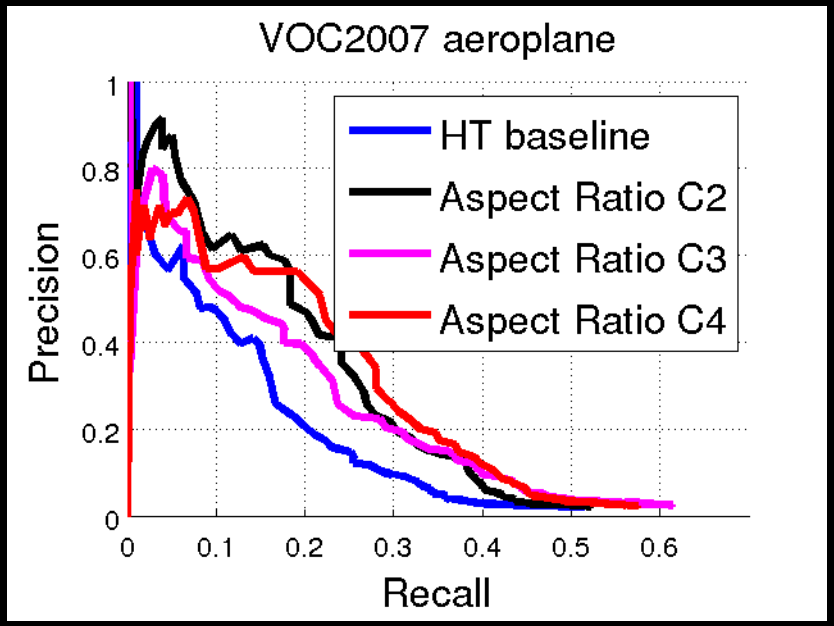


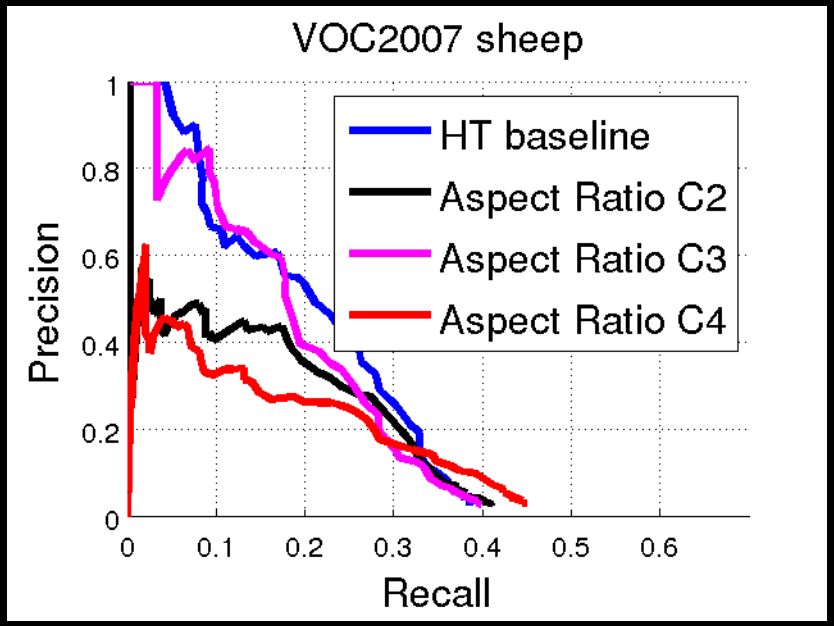
Learning or Clustering?



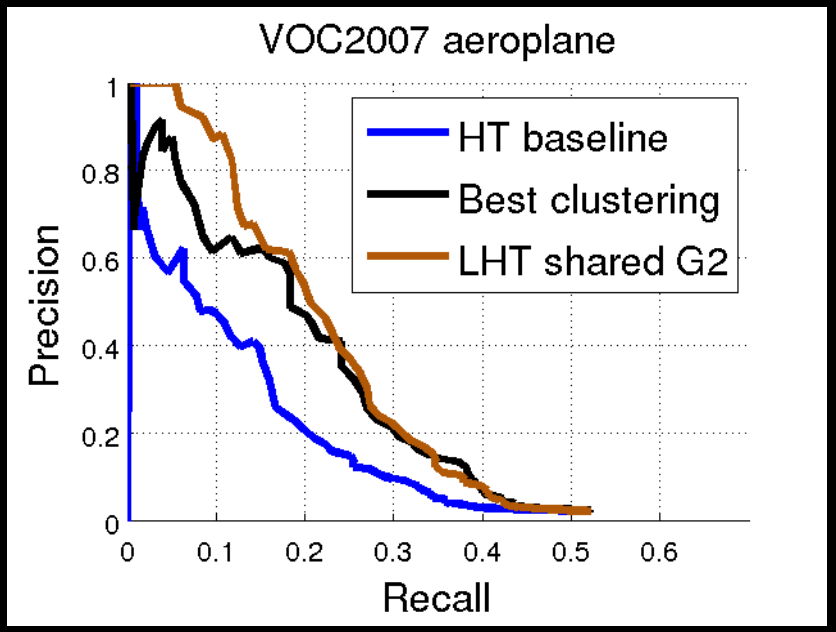


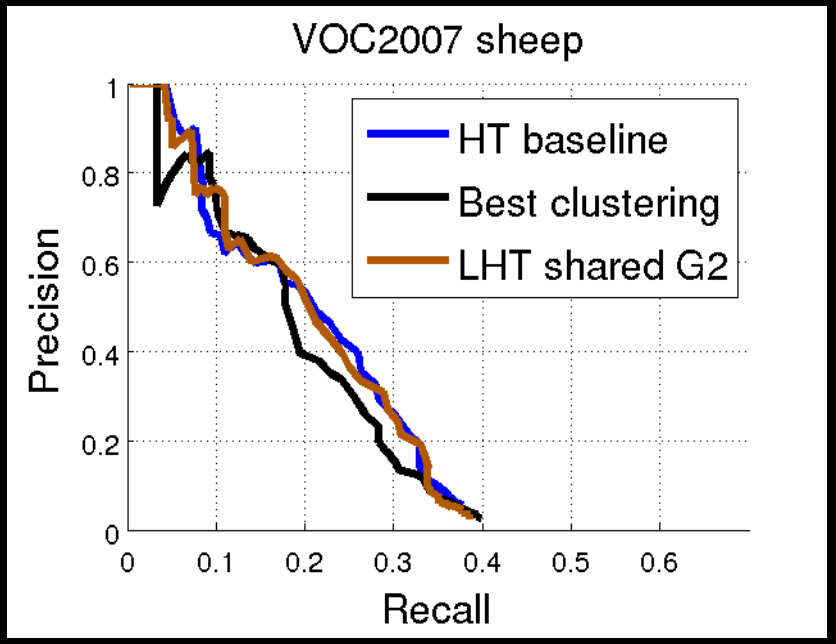
Learning or Clustering?



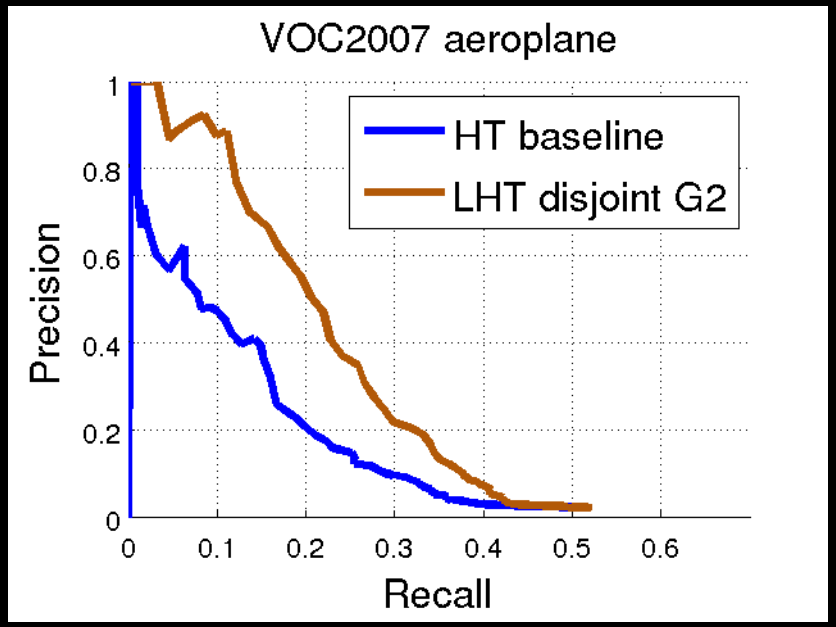


Learning or Clustering?

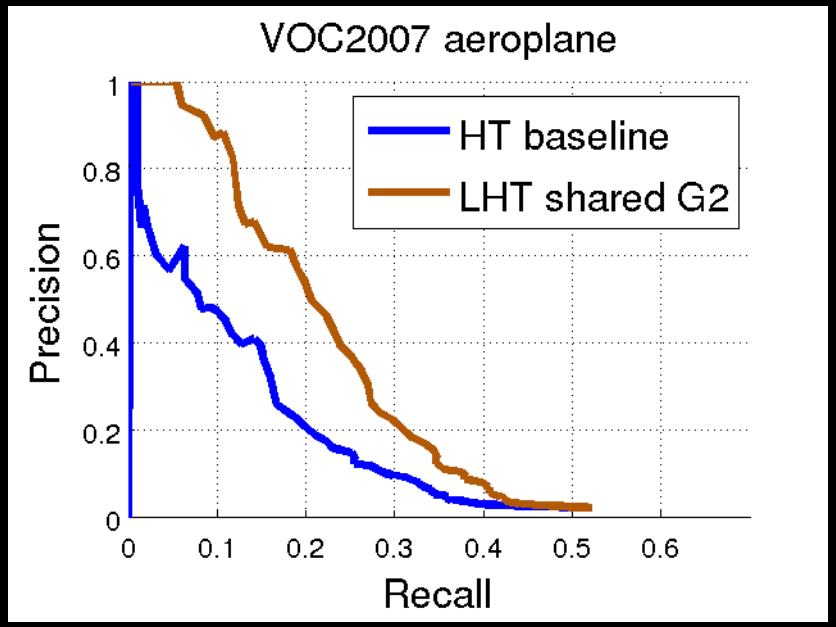




Disjoint or Shared Groups?

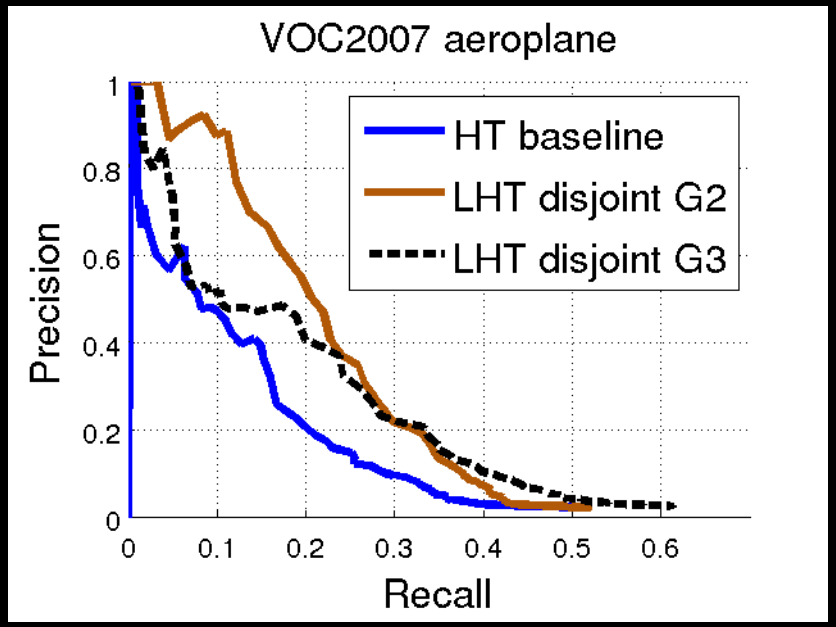


Disjoint Groups

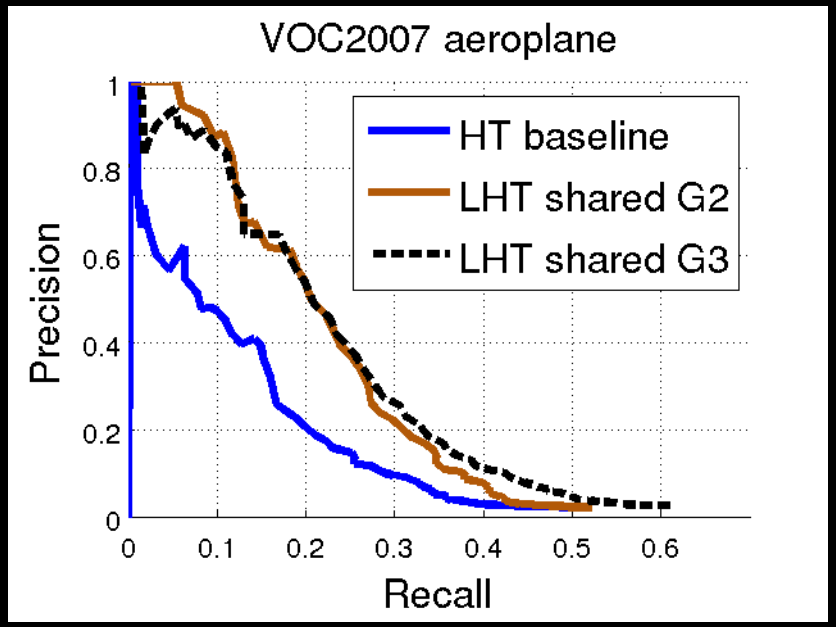


Shared Groups

Disjoint or Shared Groups?



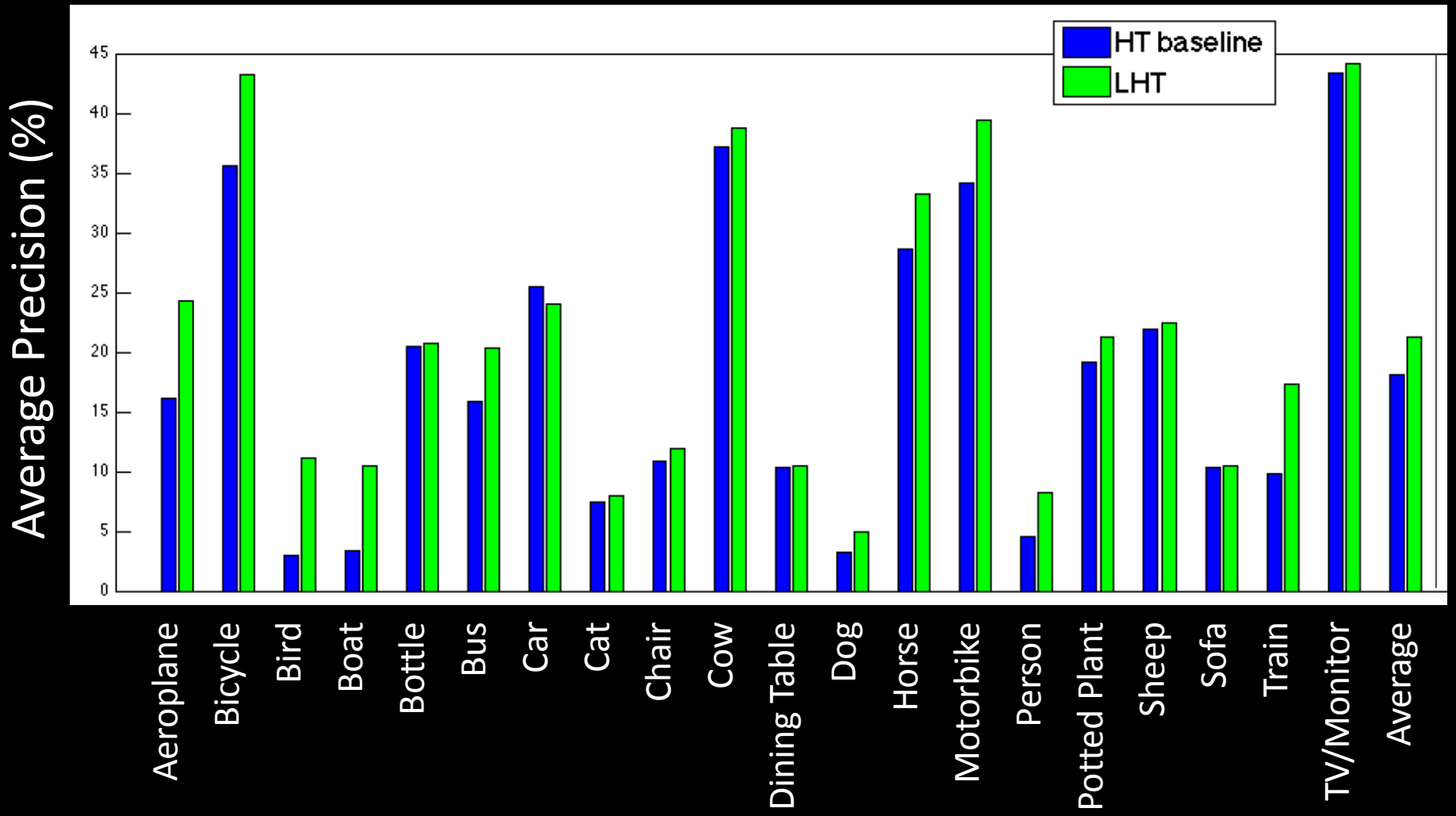
Disjoint Groups



Shared Groups

Overall Results

PASCAL VOC 2007



Contributions

- Introduced Latent Hough Transform to enforce consistency of the votes
- Discriminative learning of the latent space for object detection
- State-of-the-art performance for voting based methods

Visualization of Groups

Ignored Examples
1st Group
2nd Group
3rd Group

1st Group







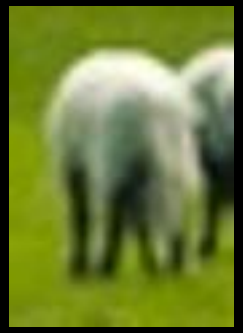
2nd Group

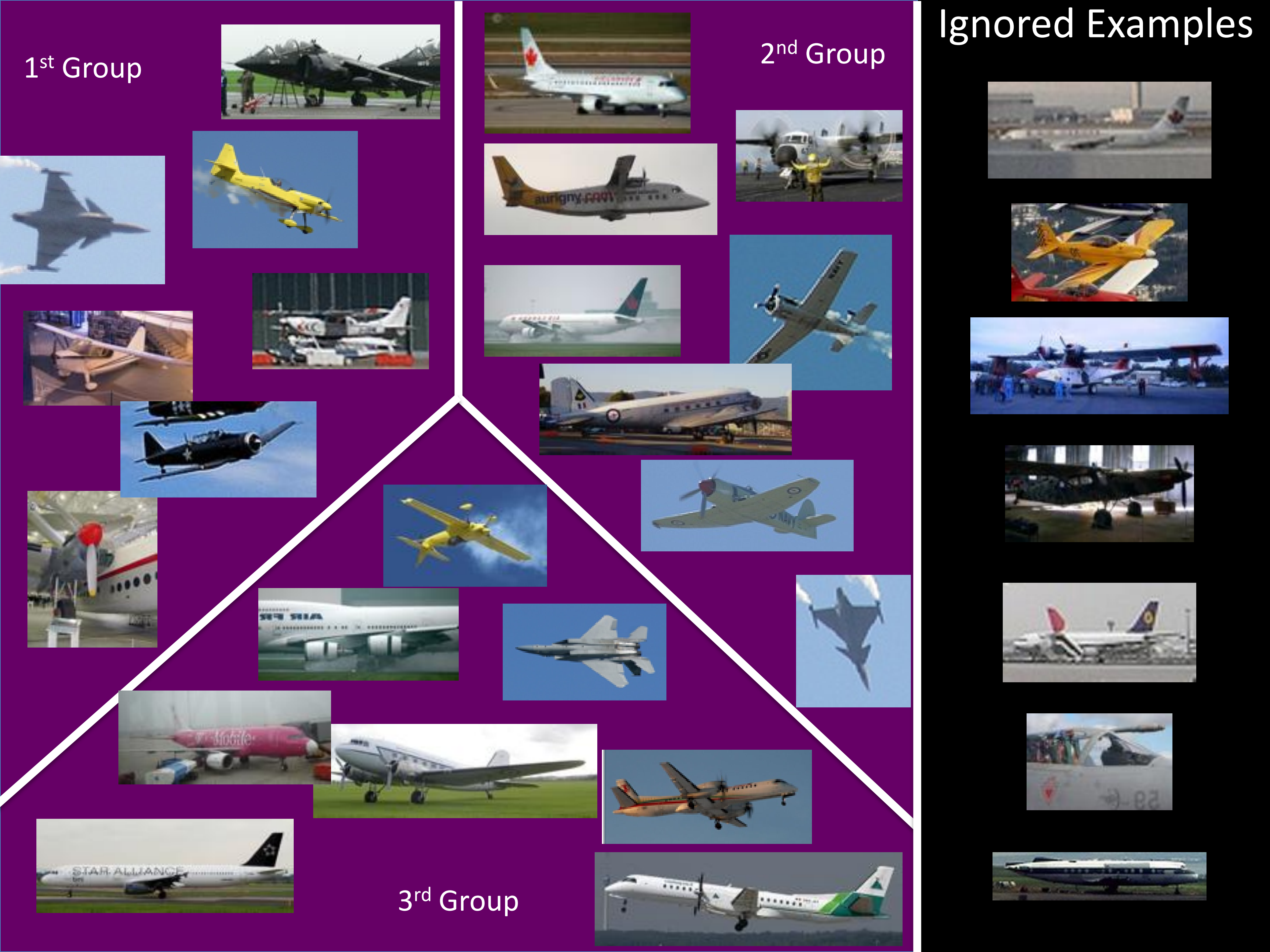






Ignored Examples



1st Group
2nd Group
3rd Group
Ignored Examples
aurigny
STAR ALLIANCE

Thank You!