Beyond bags of features: Adding spatial information

- Global spatial layout: spatial pyramid matching
- Spatial weighting the features

Spatial pyramid matching

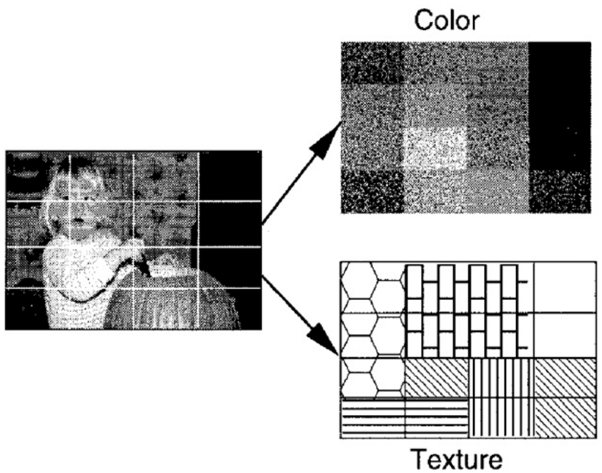
- Add spatial information to the bag-of-features
- Perform matching in 2D image space

[Lazebnik, Schmid & Ponce, CVPR 2006]

Related work

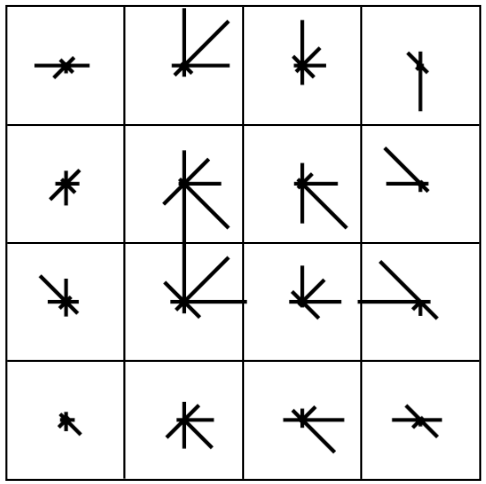
Similar approaches:

- Subblock description [Szummer & Picard, 1997]
- SIFT [Lowe, 1999]
- GIST [Torralba et al., 2003]



Szummer & Picard (1997)

SIFT

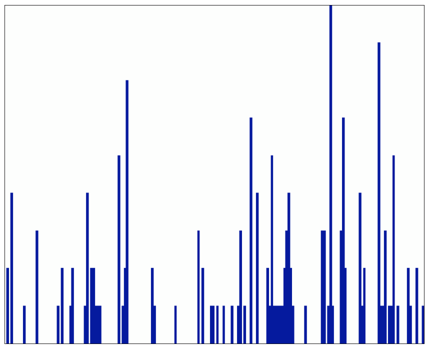

Lowe (1999, 2004)

Gist

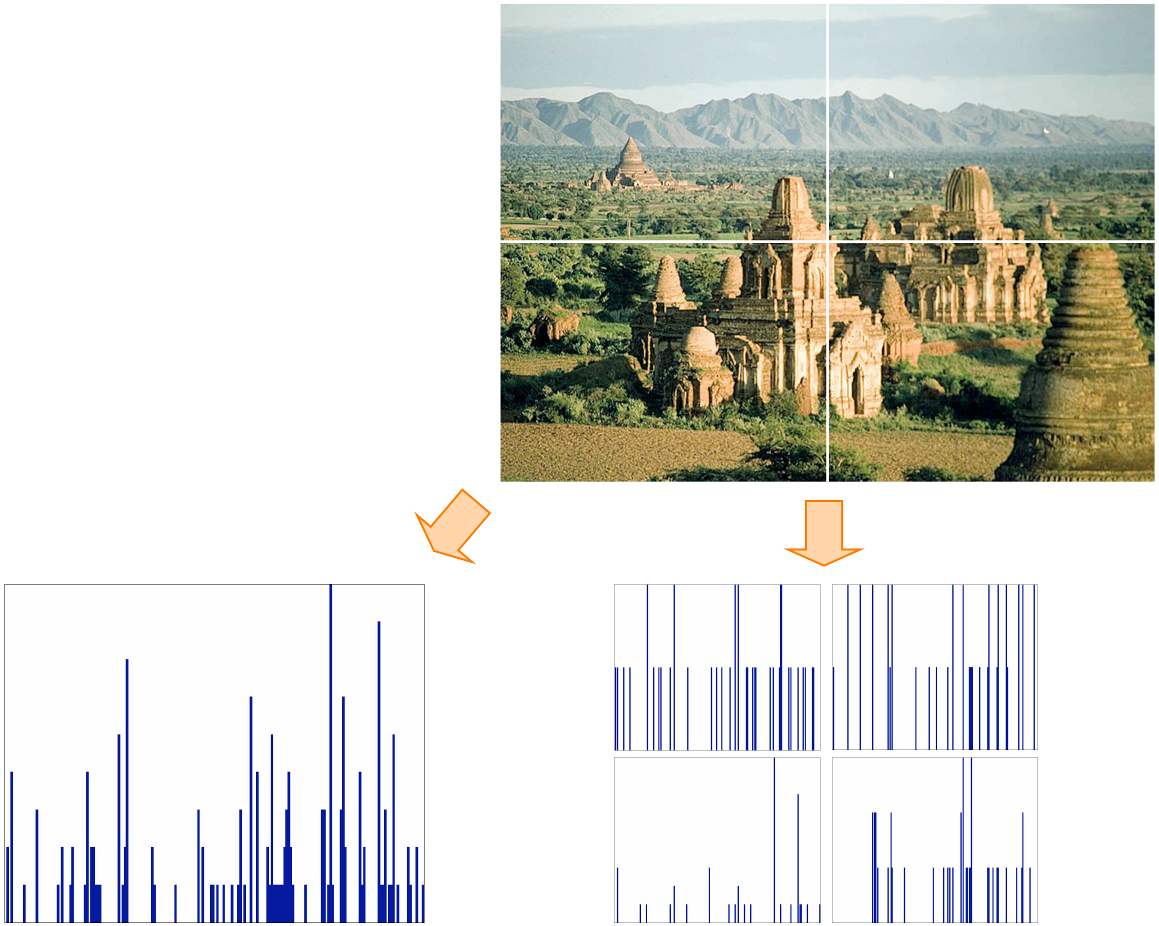

Torralba et al. (2003)

Spatial pyramid representation

Locally orderless representation at several levels of spatial resolution

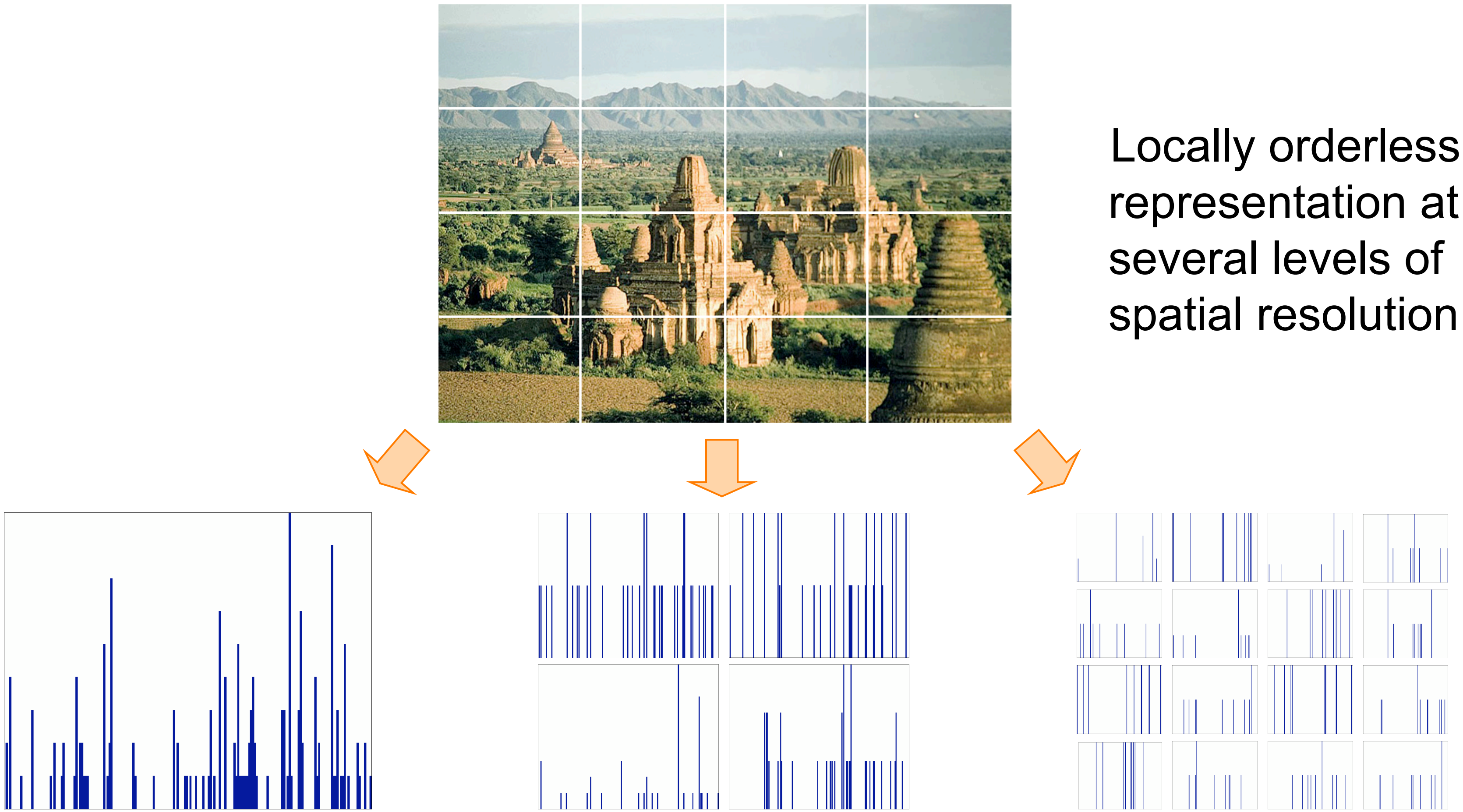


Spatial pyramid representation



Locally orderless representation at several levels of spatial resolution

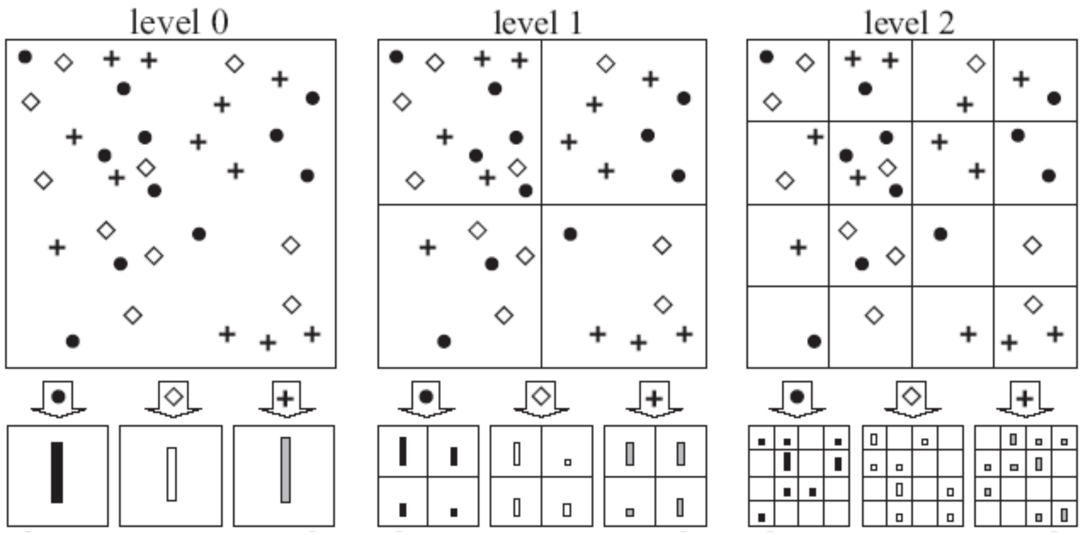
Spatial pyramid representation



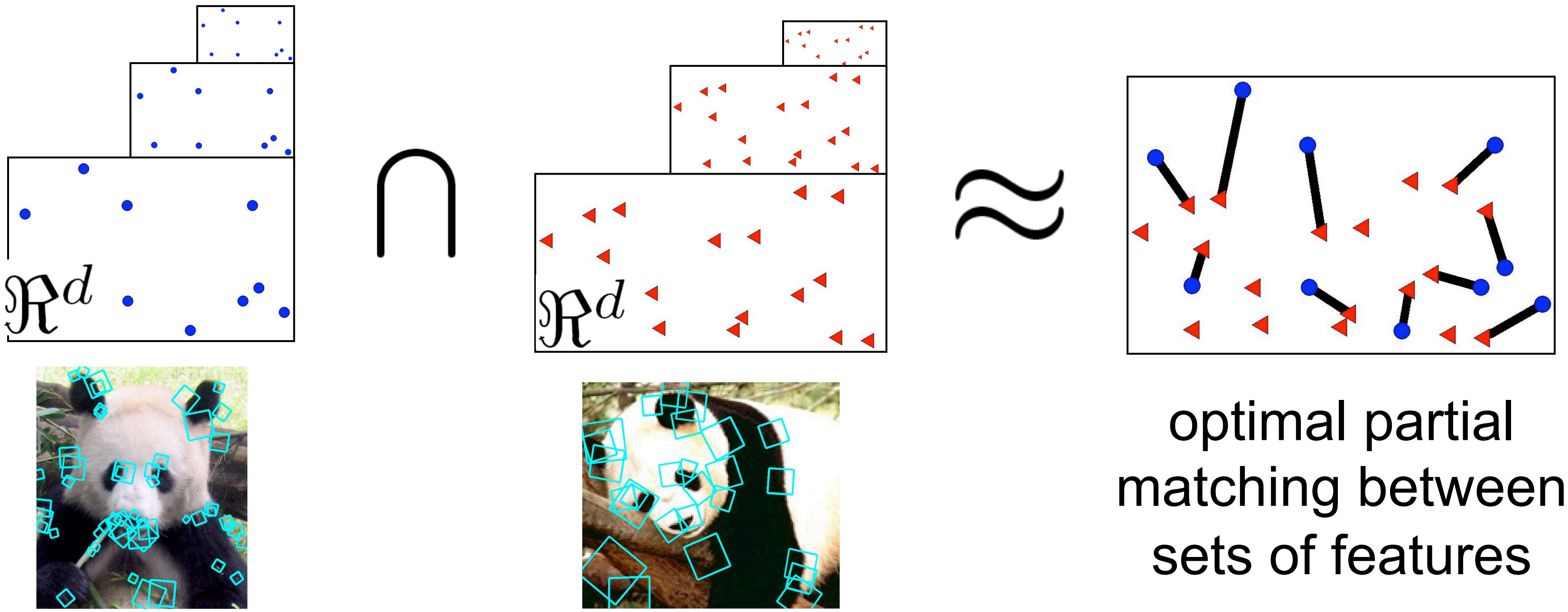
Locally orderless representation at several levels of spatial resolution

Spatial pyramid matching

- Combination of spatial levels with pyramid match kernel
  [Grauman & Darell'05]



Pyramid match kernel [Grauman & Darell'05]



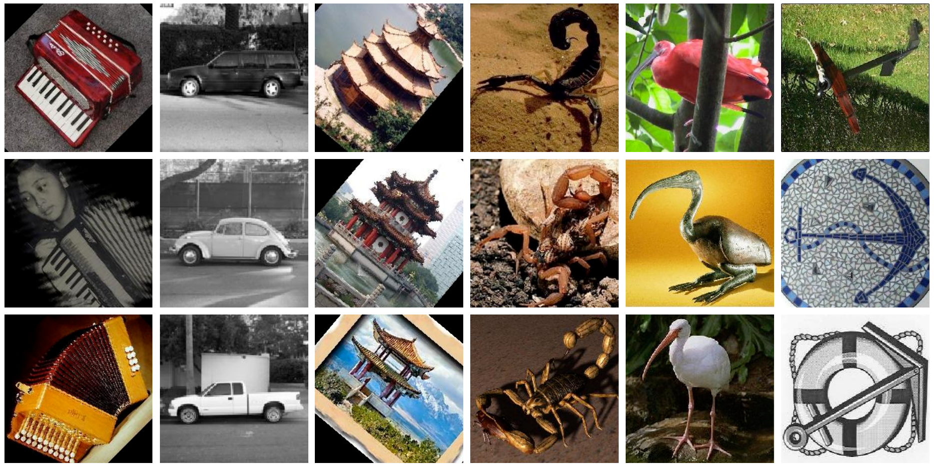
optimal partial matching between sets of features

Scene classification

L	Single-level	Pyramid
0(1x1)	72.2±0.6	
1(2x2)	77.9±0.6	79.0 ±0.5
2(4x4)	79.4±0.3	*81.1 ±0.3*
3(8x8)	77.2±0.4	80.7 ±0.3

Retrieval examples

Category classification – CalTech101

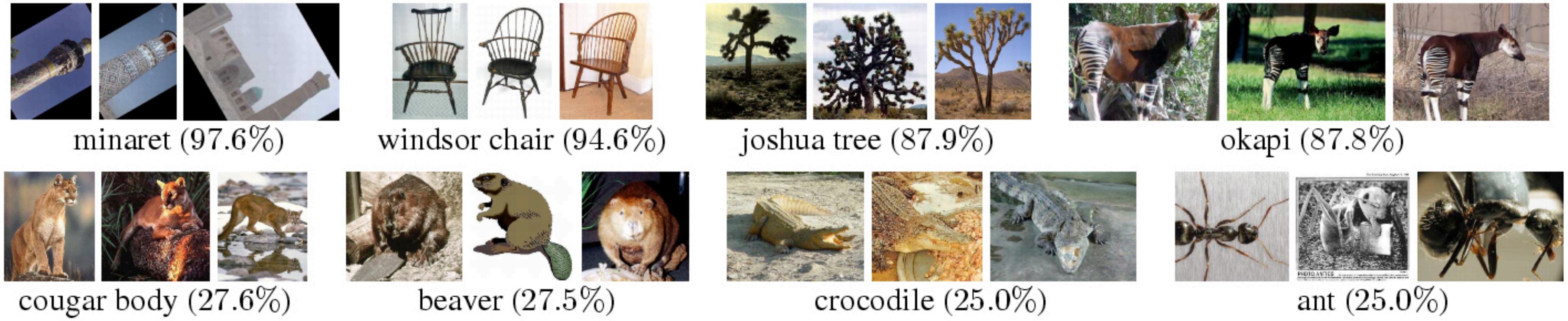


L	Single-level	Pyramid
0(1x1)	41.2±1.2	
1(2x2)	55.9±0.9	57.0 ±0.8
2(4x4)	63.6±0.9	*64.6 ±0.8*
3(8x8)	60.3±0.9	64.6 ±0.7

Bag-of-features approach by Zhang et al.'07: 54 %

CalTech101

Easiest and hardest classes



- Sources of difficulty:
  - Lack of texture
  - Camouflage
  - Thin, articulated limbs
  - Highly deformable shape

Discussion

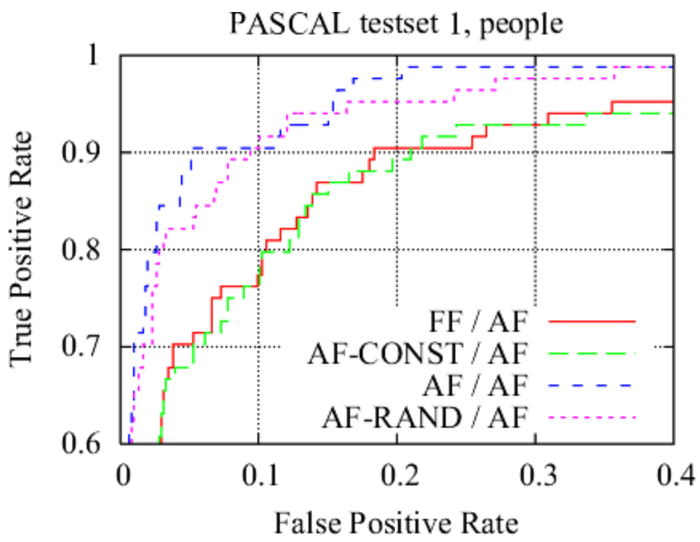
- Summary
  - Spatial pyramid representation: appearance of local image patches + coarse global position information
  - Substantial improvement over bag of features
  - Depends on the similarity of image layout
- Extensions
  - Integrating different types of features, learning weights, use of different grids [Zhang'07, Bosch & Zisserman'07, Varma et al.'07, Marszalek et al.'07]
  - Flexible, object-centered grid

Overview

- Global spatial layout: spatial pyramid matching
- *Spatial weighting the features*

Motivation

- Evaluating the influence of background features [J. Zhang, M. Marszalek, S. Lazebnik & C. Schmid, IJCV'07]
  - Train and test on different combinations of foreground and background by separating features based on bounding boxes



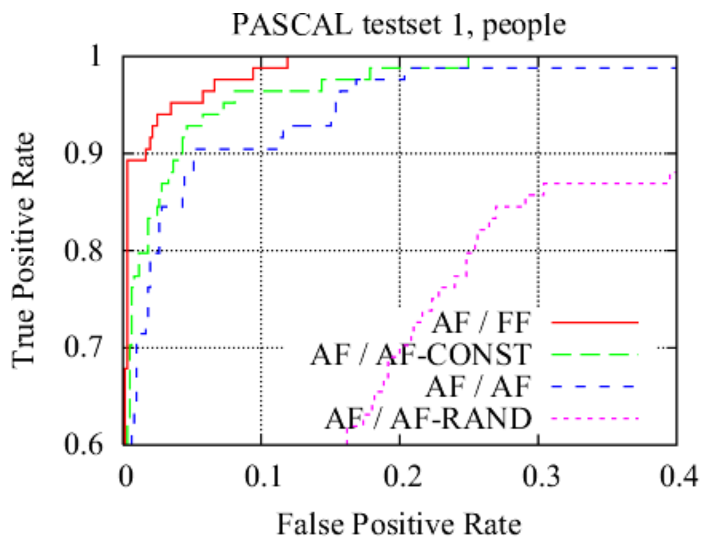
Training: different combinations foreground + background features

Testing: original test set

Best results when training with "harder" dataset (with background)

Motivation

- Evaluating the influence of background features [J. Zhang, M. Marszalek, S. Lazebnik & C. Schmid, IJCV'07]
  - Train and test on different combinations of foreground and background by separating features based on bounding boxes



Training: original training set

Testing: different combinations foreground + background features

Best results when testing with foreground features only

Approach

- Better to train on a “harder” dataset with background clutter and test on an easier one without background clutter

- Spatial weighting for bag-of-features [Marszalek & Schmid, CVPR’06]
  – weight features by the likelihood of belonging to the object
  – determine likelihood based on shape masks

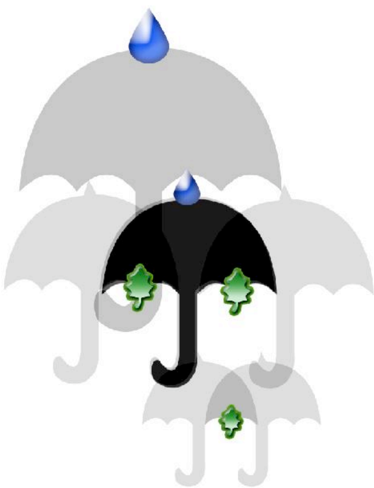
Masks for spatial weighting

For each test feature:

- Select closest training features + corresponding masks
(training requires segmented images or bounding boxes)

- Align mask based on local co-ordinates system
(transformation between training and test co-ordinate systems)

Sum masks weighted by matching distance



three features agree on object localization,
the object has higher weights

Weight histogram features with the strength of the final mask

Example masks for spatial weighting

Classification for PASCAL dataset

	Zhang et al.	Spatial weighting	Gain
bikes	74.8	76.8	+2.0
cars	75.8	76.8	+1.0
motorbikes	78.8	79.3	+0.5
people	76.9	77.9	+1.0

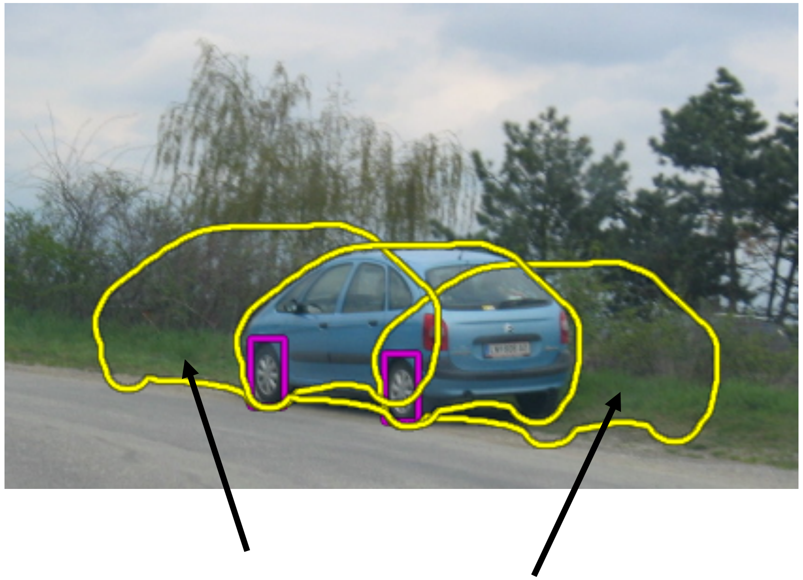
Equal error rates for PASCAL test set 2

Extension to localization

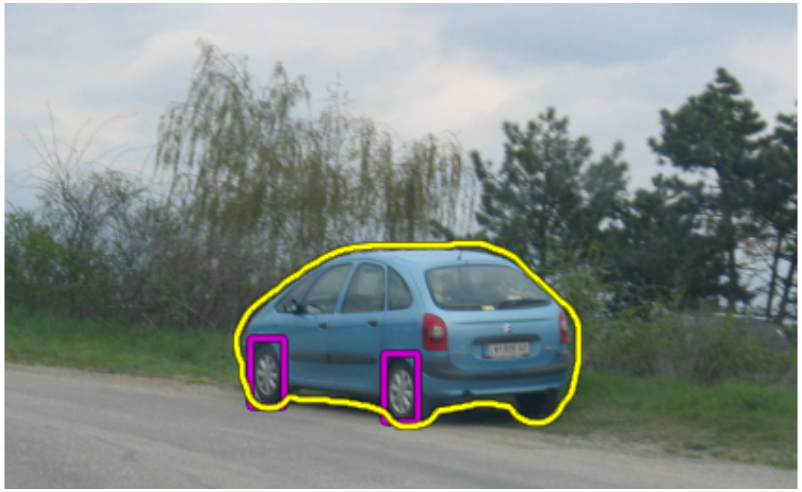
- Cast hypothesis
  - Aligning the mask based on matching features
- Evaluate each hypothesis
  - SVM for local features
- Merge hypothesis to produce localization decisions
  - Online clustering of similar hypothesis, rejection of weak ones

[Marszalek & Schmid, CVPR 2007]

Illustration of hypothesis evaluation



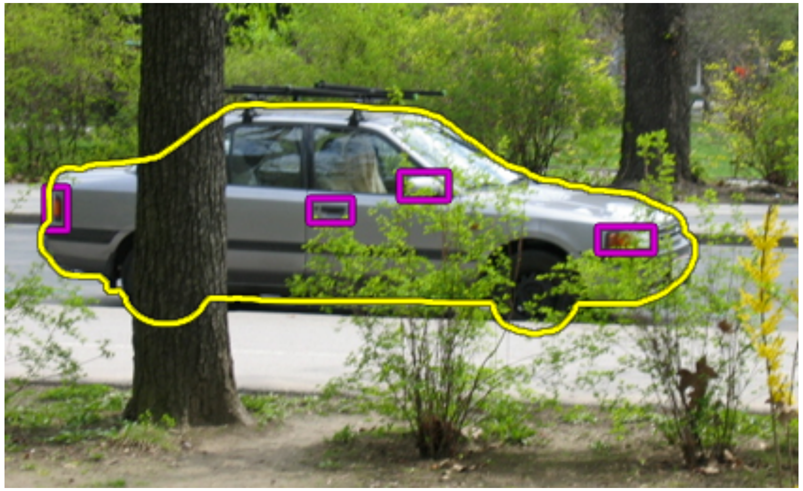
False hypotheses due to the ambiguities of the wheels



Eliminated after the evaluation

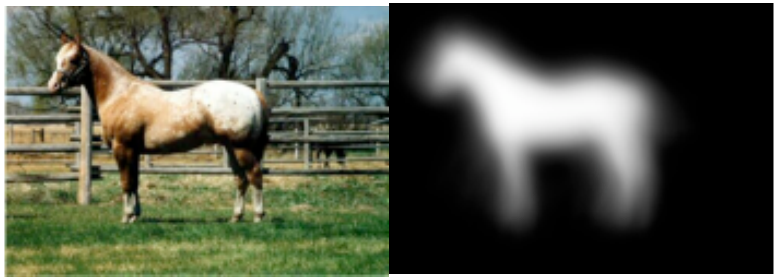
Illustration of hypotheses merging

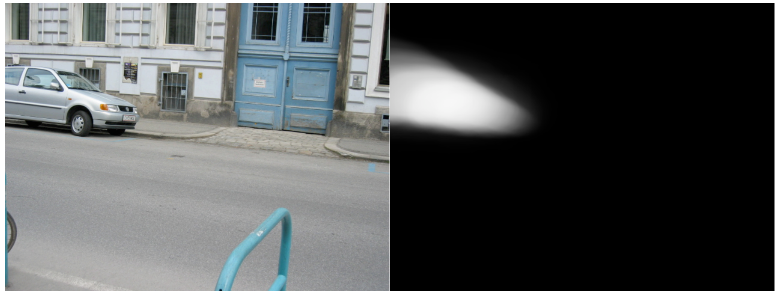
Weak classifier response
due to occlusion



Merging of evidence based on
consistent object features

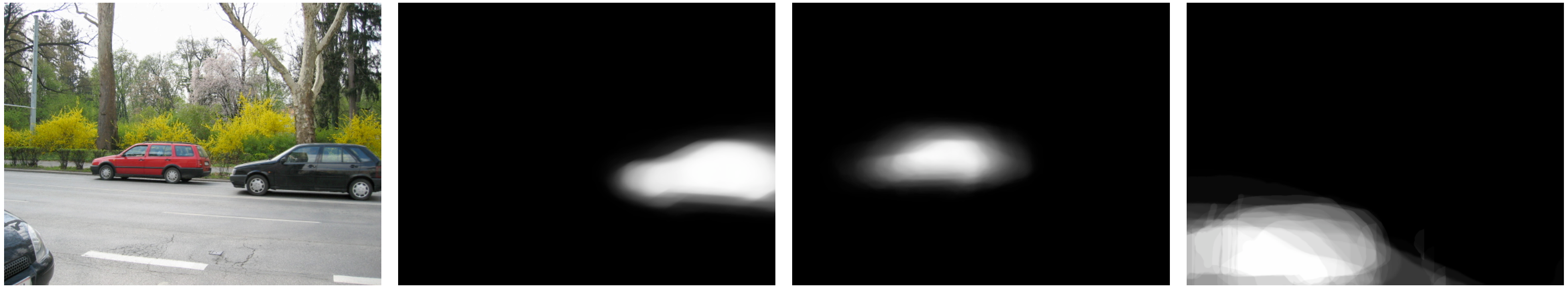
Localization results





Localization result

Illustration of subsequent hypotheses



Confidence value	**1103.1**	**561.8**	**4.9**

object class	cars	people	bicycles
no hypothesis evaluation	40.40%	28.40%	46.60%
no evidence collection	50.30%	40.30%	48.90%
our full framework	**53.80%**	**44.10%**	**61.80%**

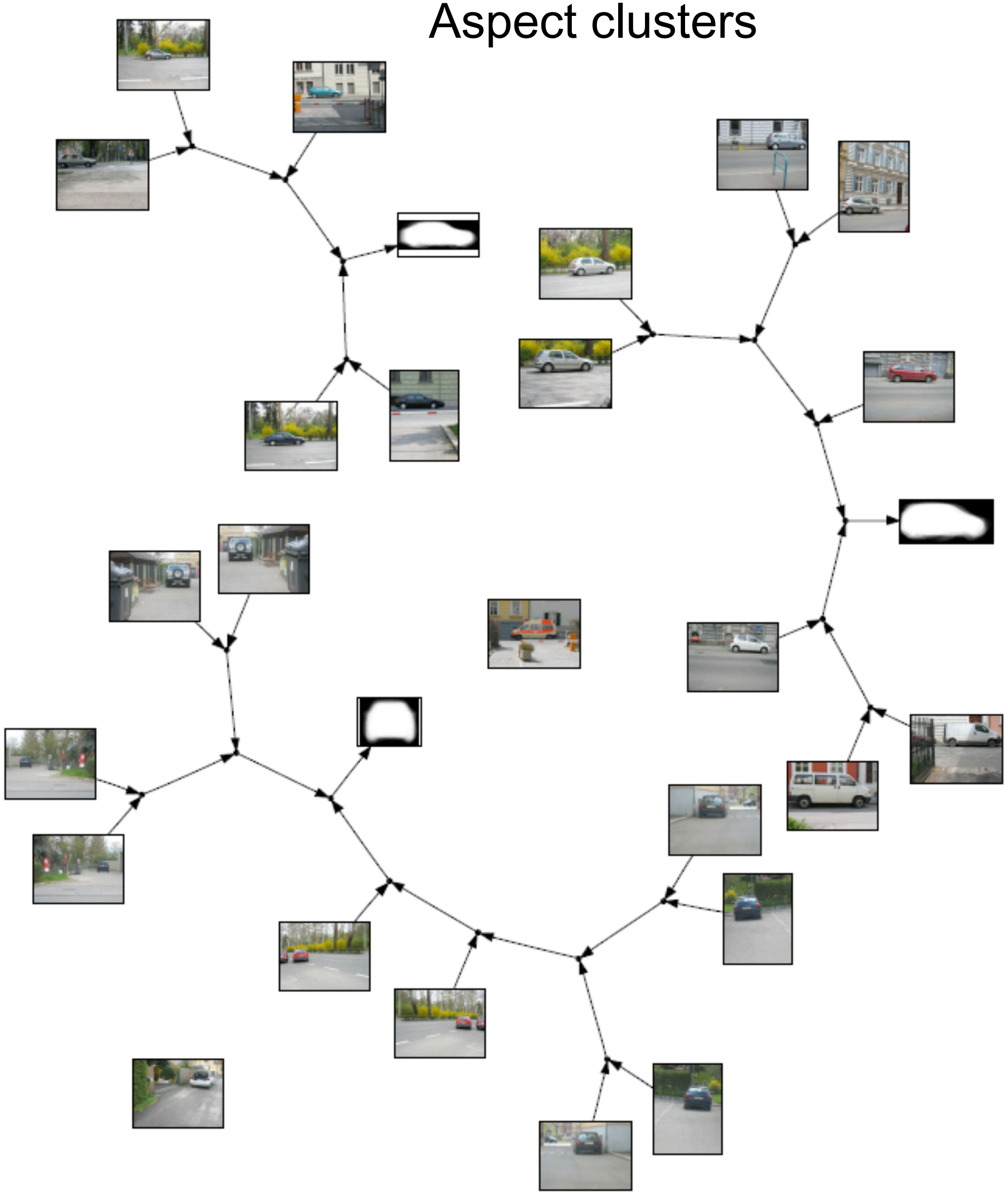
Comparison to state-of-the-art




Comparison with [Shotton et al. ICCV'05]
- use their images, search at a single scale
- improved performance over them, and:
- no use of shape-based features
- can detect objects at multiple scales

Shotton	92.10%
Our framework (no singleton pruning)	**94.60%**
Our framework (with)	**94.6**

Aspect clusters



Discussion

- Including spatial information improves results
- Importance of flexible modeling of spatial information
  - coarse global position information
  - object based models
- Extensions
  - Hierarchical organization of the objects/aspects