Spatial Weighting for Bag-of-Features

Marcin Marszalek (marszale@inrialpes.fr) **Cordelia Schmid** (schmid@inrialpes.fr)

INRIA Rhône-Alpes, LEAR - GRAVIR, Montbonnot, France

1. Motivation

We extend the state-of-the art image classification method of Zhang et al. (to appear in IJCV):

- interest points (Harris, LoG)

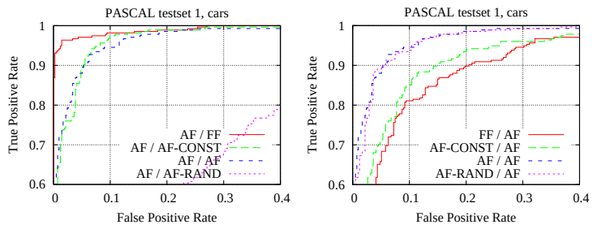
- local descriptors (SIFT)
- bag of features (histograms of visual words H_i)
- SVMs with χ^2 kernel for classification

$$K(H_i, H_j) = e^{-\frac{1}{A}D(H_i,H_j)}$$

$$D(H_i, H_j) = \frac{1}{2}\sum_{n=1}^{N}\frac{(h_{in} - h_{jn})^2}{h_{in} + h_{jn}}$$

We exploit some of the paper's conclusions:

- background removal improves the classification results
- one should not train on too easy examples



AF	all features (foreground and background features)
FF	foreground features (background removed)
FF-CONST	constant scene background (foreground features on mostly static natural background features)
AF-RAND	randomized background (foreground features with background features of a random image)

2. Spatial Weighting

Overview:

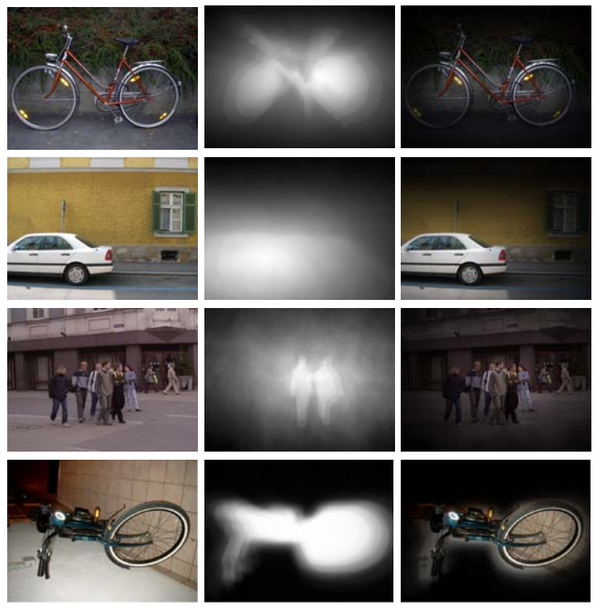
- we perform training in strongly supervised fashion
- we employ spatial relationships between features
- we use those relations to reduce background clutter

Goal: Features that agree on the position and shape of the object should have their weights increased, while background features should be suppressed.

Method: We compute an approximate segmentation mask and reweight the features in the histogram.

Algorithm:

- for each test feature we look for the most similar training features and retrieve the associated segmentations
- we normalize (translate, rescale, etc.) the segmentation masks to compensate for the descriptor invariance
- we stack the masks and compute a weighted sum, which considers the similarity of matched descriptors
- we reweight the foreground (located within the approximated segmentation mask) and background features (located outside)
- we compute the new histogram for classification



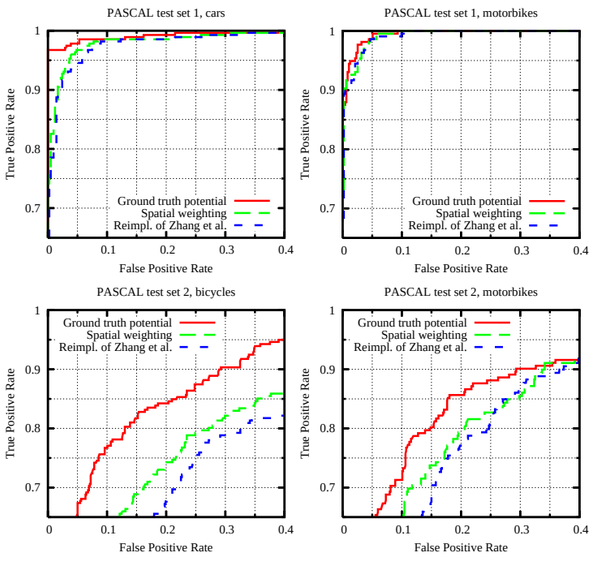
Related work: Leibe and Schiele (ECCV'04):

- we do not use the Hough transform, we support arbitrary object shapes without simplifying them
- patch-based segmentation does not work well with sparse image representation; instead of dense sampling we employ full segmentation masks

3. Experimental results

- we evaluate on the demanding, natural-scene PASCAL challenge 2005 dataset

ROC EER		Reimpl. of Zhang et al. (HS+LS)-SIFT	Spatial Weighting (HS+LS)-SIFT	Gain	Winner
test set 1	bikes	92.1	92.1		93.0
	cars	94.5	96.0	+1.5	96.1
	motorbikes	96.3	96.3		97.7
	people	91.7	92.9	+1.2	91.7
test set 2	bikes	74.8	76.8	+2.0	72.8
	cars	75.8	76.8	+1.0	72.0
	motorbikes	78.8	79.3	+0.5	79.8
	people	76.9	77.9	+1.0	71.9



4. Extensions

- iterative segmentation mask refinement is possible

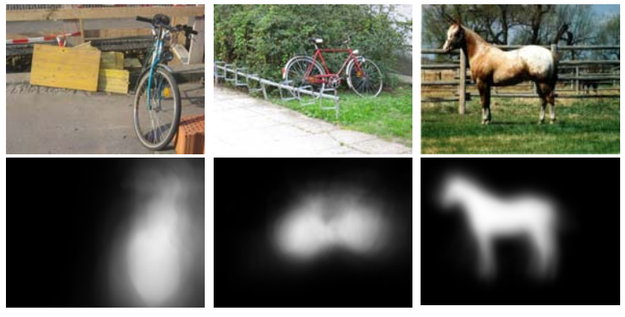
- segmentation masks can be used to localize objects

	Spatial Weighting HS-SIFT	Spatial Weighting LS-SIFT	Opelt
bikes	78.7	82.7	76.7
cars	62.7	68.0	55.3
people	83.3	71.3	48.0

5. Future work

- support the localization of multiple object instances

- improve the produced segmentation masks



6. Summary

- we extend the bag-of-features representation to include spatial relationships between features
- we improve the image classification accuracy
- we localize object instances providing information about their position and state

M. Marszalek is supported by a grant from the European Community under the Marie-Curie project VISITOR. This work was supported by the European Network of Excellence PASCAL.