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]
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 \cite{Grauman & Darell'05}

\[ \mathbb{R}^d \bigcap \mathbb{R}^d \approx \text{optimal partial matching between sets of features} \]
Scene classification

<table>
<thead>
<tr>
<th>L</th>
<th>Single-level</th>
<th>Pyramid</th>
</tr>
</thead>
<tbody>
<tr>
<td>0(1x1)</td>
<td>72.2±0.6</td>
<td></td>
</tr>
<tr>
<td>1(2x2)</td>
<td>77.9±0.6</td>
<td>79.0±0.5</td>
</tr>
<tr>
<td>2(4x4)</td>
<td>79.4±0.3</td>
<td>81.1±0.3</td>
</tr>
<tr>
<td>3(8x8)</td>
<td>77.2±0.4</td>
<td>80.7±0.3</td>
</tr>
</tbody>
</table>
Category classification – CalTech101

<table>
<thead>
<tr>
<th>L</th>
<th>Single-level</th>
<th>Pyramid</th>
</tr>
</thead>
<tbody>
<tr>
<td>0(1x1)</td>
<td>41.2±1.2</td>
<td></td>
</tr>
<tr>
<td>1(2x2)</td>
<td>55.9±0.9</td>
<td>57.0 ±0.8</td>
</tr>
<tr>
<td>2(4x4)</td>
<td>63.6±0.9</td>
<td>64.6 ±0.8</td>
</tr>
<tr>
<td>3(8x8)</td>
<td>60.3±0.9</td>
<td>64.6 ±0.7</td>
</tr>
</tbody>
</table>

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.

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

<table>
<thead>
<tr>
<th></th>
<th>Zhang et al.</th>
<th>Spatial weighting</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>bikes</td>
<td>74.8</td>
<td>76.8</td>
<td>+2.0</td>
</tr>
<tr>
<td>cars</td>
<td>75.8</td>
<td>76.8</td>
<td>+1.0</td>
</tr>
<tr>
<td>motorbikes</td>
<td>78.8</td>
<td>79.3</td>
<td>+0.5</td>
</tr>
<tr>
<td>people</td>
<td>76.9</td>
<td>77.9</td>
<td>+1.0</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>object class</th>
<th>cars</th>
<th>people</th>
<th>bicycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>no hypothesis evaluation</td>
<td>40.40%</td>
<td>28.40%</td>
<td>46.60%</td>
</tr>
<tr>
<td>no evidence collection</td>
<td>50.30%</td>
<td>40.30%</td>
<td>48.90%</td>
</tr>
<tr>
<td>our full framework</td>
<td>53.80%</td>
<td>44.10%</td>
<td>61.80%</td>
</tr>
</tbody>
</table>
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

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Shotton</td>
<td>92.10%</td>
</tr>
<tr>
<td>Our framework (no singleton pruning)</td>
<td>94.60%</td>
</tr>
<tr>
<td>Our framework (with)</td>
<td><strong>94.6</strong></td>
</tr>
</tbody>
</table>
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