BİL-722 Advanced Topics in Computer Vision

Presentation: Çağdaş Baş
Paper: Randomized Spatial Partition for Scene Recognition
Authors: Yuning Jiang, Junsong Yuan, and Gang Yu
RANDOMIZED SPATIAL PARTITION FOR SCENE RECOGNITION

Images from: SUN Dataset
Related Work

- Behmo et. al. (ECCV2010)
  - Feature points are independent

- Zhang et. al. (ACM2009)
  - Only relative positions are considered

- Lazebnik et. al. (CVPR2006)
  - Fixed partitioning
Spatial Pyramid Matching
Lazebnik et. Al. (CVPR2006)

- Build pyramid in image space, quantize feature space
**Pyramid matching**

Indyk & Thaper (2003), Grauman & Darrell (2005)

Find maximum-weight matching (weight is inversely proportional to distance)

Original images:

Feature histograms:
Level 3

Level 2

Level 1

Level 0
Limitations of SPM

- Different spatial layouts for similar scenes (a)
- Intra-class variations due to scale, viewpoint and rotation (b)

Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu
Limitations of Spm Images from SUN dataset
Randomized Patterns

- Instead of making fixed partitions, randomize partition for each class

Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu
Choosing Right Pattern: Optimal Pattern Selection

- Find best pattern for each level with lowest error with SVM classification:

\[
err_\theta^c = \sum_{c_v = c} I(f_\theta^c(I_v) \neq c) + \sum_{c_v \neq c} I(f_\theta^c(I_v) = c),
\]

Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu
Choosing Right Pattern: Boosting

1. **For all** \( \theta \in \Theta \):
   - Randomly sample a subset \( \Phi_\theta \subset \Phi \), and represent the images in \( \Phi_\theta \) in pattern \( \theta \).
   - Train a multi-class classifier \( f_\theta(\cdot) \) on the random subset \( \Phi_\theta \) using SVM.

2. Initialize the weight \( w_i = \frac{1}{CN_{c_i}} \) for each images \( I_i \), where \( N_{c_i} \) is the number of the images with label \( c_i \); the current iteration number \( j = 0 \); the current accuracy \( \sigma^{(0)} = 0 \).

3. **While** \((\sigma^{(j)} < \sigma_{\text{target}})\):
   - \( \forall i = 1, \ldots, N, w_i \leftarrow \frac{w_i}{\sum_{i=1}^{N} w_i}; j \leftarrow j + 1 \).
   - \( \forall \theta \in \Theta \), calculate its classification error on \( \Phi \): \( err_\theta = \sum_{i \in \Phi} w_i \cdot I(f_\theta(I_i) \neq c_i) \).
   - Select the pattern \( \theta^{(j)} \) with minimum error \( err^{(j)} \), and then calculate the weight for \( \theta^{(j)} \) as:
     \[
     \alpha^{(j)} = \log \frac{1 - err^{(j)}}{err^{(j)}} + \log(C - 1).
     \]
   - \( \forall i = 1, \ldots, N, w_i \leftarrow w_i \cdot \exp(\alpha^{(j)} \cdot I(f_{\theta^{(j)}}(I_i) \neq c_i)) \).
   - Generate the strong classifier as:
     \[
     F(I) = \arg \max_c \sum_{m=1}^{j} \alpha^{(m)} \cdot I(f_{\theta^{(m)}}(I) = c),
     \]
   and calculate its classification accuracy on \( \Phi \): \( \sigma^{(j)} = \sum_{I_i \in \Phi} I(F(I_i) = c_i)/N \).

Algorithm Credit: Yuning Jiang, Junsong Yuan, Gang Yu
Experimental Results

- Error rates for 15-Scene dataset

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$l = 0$</td>
<td>74.8%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$l = 1$</td>
<td>78.8%</td>
<td>80.7%</td>
<td>86.4%</td>
</tr>
<tr>
<td>$l = 2$</td>
<td>79.7%</td>
<td>82.6%</td>
<td>87.1%</td>
</tr>
<tr>
<td>$l = 0, 1, 2$</td>
<td>81.4%</td>
<td>83.9%</td>
<td>87.2%</td>
</tr>
</tbody>
</table>

Table 1. Comparison with SPM at different levels on the 15-scene dataset.
Experimental Results
15-scene Dataset

- 15-scene dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Avg. Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPM + SIFT with 400 clusters [11]</td>
<td>81.4%</td>
</tr>
<tr>
<td>SPM + SIFT with 400 concepts [28]</td>
<td>83.3%</td>
</tr>
<tr>
<td>DSP + SIFT with 1000 clusters [23]</td>
<td>80.7%</td>
</tr>
<tr>
<td>SP-pLSA + SIFT with 1200 topics [29]</td>
<td>83.7%</td>
</tr>
<tr>
<td>CENTRIST + RBF-SVM [12]</td>
<td>83.9%</td>
</tr>
<tr>
<td>CENTRIST + LCC + Boosting [27]</td>
<td>87.8%</td>
</tr>
<tr>
<td>RSP + Optimal Selection</td>
<td>83.9%</td>
</tr>
<tr>
<td>RSP + Boosting ($</td>
<td>\Phi_\theta</td>
</tr>
<tr>
<td>RSP + Boosting ($</td>
<td>\Phi_\theta</td>
</tr>
</tbody>
</table>

Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu
Experimental Results
15-scene Dataset – Cont.

Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu
Mis-Classifications
For 15-scene Dataset

- Images for coast-open-country from 15-scene dataset
- Images for living-room-bedroom from 15-scene dataset
- Images for inside-city-industrial from 15-scene dataset
## Experimental Results
### 8-Scene Dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Avg. Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scene Model + SIFT [14]</td>
<td>≈ 60%</td>
</tr>
<tr>
<td>Scene Model + Object Model + SIFT [14]</td>
<td>73.4%</td>
</tr>
<tr>
<td>PACT + RBF-SVM [12]</td>
<td>78.2%</td>
</tr>
<tr>
<td>SPM + RBF-SVM</td>
<td>74.0%</td>
</tr>
<tr>
<td>RSP + Optimal Selection</td>
<td>77.9%</td>
</tr>
<tr>
<td>RSP + Boosting</td>
<td><strong>79.6%</strong></td>
</tr>
</tbody>
</table>

Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu
Experimental Results

21-land-use Dataset

- Aerial images of scenes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc.</td>
<td>71.9</td>
<td>74.0</td>
<td>73.1</td>
<td>76.1</td>
<td>77.3</td>
<td>75.5</td>
<td><strong>77.8</strong></td>
</tr>
</tbody>
</table>

Figure Credit: Yuning Jiang, Junsong Yuan, Gang Yu