Efficient Kernels Couple Visual Words Through Categorization Opponency

I. Alexiou
A.A. Bharath
Overview of Talk

- Introduction – categorization from tunable word pairs
- Related Work
- Visual Structure Pairing
- Experiments
- Results
- Conclusions & Future Work
Related Work


Typical Hierarchical Example

Simplified to a single node

Replace leaves with visual words

Levels of Hierarchy → Spatial Location → Hidden Units

Empirically, it is known that different words co-occur in images. In order to detect co-occurrences we formulate this idea to a probabilistic approach.

$$P(VW_i \cap VW_j) = P(VW_j|VW_i)P(VW_i)$$

This joint probability express the likelihood two words co-occur.

We want to compute the likelihood of $VW_j$ given $VW_i$ has occurred.

Do we have any prior knowledge of this word?

$$G_i = - \ln P(VW_i)$$

$K_h(w_i, w_j) = \{ w_1, w_2, \ldots, w_N \}$
Categorical Opponency

• Properties antagonistic among the classes
• Pairs that are unique to a class or have high likelihood for that class.
• If this does not exist then there is no discrimination among the categories.
Codebook of Paired Words

These ratios are ranked per category and then gathered to create a codebook of pairs.
Another Interpretation


\[ K_h(w_i, w_j) = P(w_j|w_i) \cdot G_i \]

\[ K_{Op}(w_i, w_j) = \max_{c_t} \frac{K_h(w_i, w_j, c_t)}{\max_{c_t' \neq c_t} [K_h(w_i, w_j, c_t') \leq \max_{c_t} K_h(w_i, w_j, c_t)]} \]
Tuning Pairing Behaviour

\[ r = \sigma_i + \sigma_j \]

\[ S_1(P_p|\phi_{w_j}^{(n)} \cap \phi_{w_i}^{(m)}) = d \]

\[ S_2(P_p|\phi_{w_j}^{(n)} \cap \phi_{w_i}^{(m)}) = \max\left\{0, \frac{d-r}{r}\right\} \]

\[ S_3(P_p|\phi_{w_j}^{(n)} \cap \phi_{w_i}^{(m)}) = \frac{1}{1 + e^{-\frac{d-\alpha \cdot r}{r}}} \]
Pair Detection

\[ B^{(p)} = \sum_{m | n=1}^{Q} \max_{n | m=1} K_{N \times M}^{(p)}(\phi_{w_i}^{(n)} = w_j, \phi_{w_j}^{(m)} = w_i) \]
Experiments

Caltech101 – 30 training examples

Keypoint-Based

Pascal VOC 2011 – 300 training examples

<table>
<thead>
<tr>
<th></th>
<th>Pyramids (10500)</th>
<th>Pairs (5000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keypoint-Based</td>
<td>56.34±1.625</td>
<td>66.42±2.07</td>
</tr>
<tr>
<td>Grid-Based</td>
<td>67.52±2.036</td>
<td>69.53±2.03</td>
</tr>
</tbody>
</table>

Caltech101 – Summary

<table>
<thead>
<tr>
<th></th>
<th>Pyramids (10500)</th>
<th>Pairs (5000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keypoint-Based</td>
<td>52.30±1.28</td>
<td>71.00±2.00</td>
</tr>
<tr>
<td>Grid-Based</td>
<td>60.16±1.56</td>
<td>73.88±2.10</td>
</tr>
</tbody>
</table>
Conclusions/Future Work

• A way of selecting and ranking word pairs based on ability to characterise.
• Can be taught to display higher discrimination
• Has been found to work with other methods of keypoint detection and descriptors.
• Future: plan to extend to CalTech256
• Extend the tree-structure approach.