Complexity-Adaptive Distance Metric for Object Proposals Generation

Yao Xiao, Cewu Lu, Efstratios Tsegounis, Yongyi Lu, Chi-Keung Tang
The Hong Kong University of Science and Technology

Object Proposals have become indispensable for object detection, providing the latter with a set of image regions where objects are likely to occur. Currently, the mainstream methods [1, 4] partition the image into hundreds of superpixels, and then group them under certain criteria to form object proposals. Typically, the distance metric computes the difference between two superpixels in terms of an aggregate measure. While well suited to low-complexity superpixels grouping, it becomes less effective in high-complexity scenarios, Figure 1. In this paper, we propose a novel distance metric for grouping two superpixel sets that is adaptive to their complexity. Our distance metric combines a "low-complexity distance" and a "high-complexity distance" making it adaptive to different complexities.

Our system adopts the grouping scheme of [4]. Initially, a number of superpixels are generated and in each iteration, two superpixel sets with the smallest distance are merged. However, different from [4], low-level features (histogram) are not propagated during superpixel merging. Our complexity-adaptive distance metric is composed of several basic distance:

\[d_{ct}(i,j) = ||h_c(i) - h_c(j)|| + ||h_t(i) - h_t(j)||\]  \hspace{1cm} (1)

**Graph Distance.** Our algorithm does not restrict grouping exclusively to local neighboring superpixels. Instead we use graph distance to regularize the grouping process to prefer spatially close superpixels:

\[D_g(m,n) = \min \{d_g(i,j) | i \in S_m, j \in S_n \} \]  \hspace{1cm} (2)

where \(S_m\) and \(S_n\) denote two superpixel sets.

**Edge cost** measures the edge responses along the common border of the segments. The edge cost for each neighboring segments is calculated by summing up the edge responses within the common border pixels and then normalized by the length of the common border. Denote the common border pixels set as \(L_{ij}\), then

\[D_e(m,n) = \begin{cases} \sum_{i \in S_m, j \in S_n} ||L_{ij}||_1 / ||L_{ij}|| & \text{if } \sum_{i \in S_m, j \in S_n} L_{ij} \neq 0 \\ 0 & \text{otherwise} \end{cases} \]  \hspace{1cm} (3)

Consider two super-pixel set \(S_m\) and \(S_n\), we define the nearest and farthest pairwise distance as,

\[D_{min}(m,n) = \min \{d_{ct}(i,j) | i \in S_m, j \in S_n \} \]  \hspace{1cm} (4)

\[D_{max}(m,n) = \max \{d_{ct}(i,j) | i \in S_m, j \in S_n \} \]  \hspace{1cm} (5)

\(D_{max}\) and \(D_{min}\) can be used to indicate respectively the low and high complexity distance of two given superpixels. A small \(D_{max}\) indicates that all the elements in the two sets are similar, meaning that they are of low complexity. Thus \(D_{max}\) suits for low-complexity region merging. In contrast, a small \(D_{min}\) means that the two sets are connected by at least two elements from the respective two superpixel sets. Therefore, \(D_{min}\) is a reasonable indicator for merging in high-complexity scenarios. Our low-complexity distance \(D_L\) and high-complexity distance \(D_H\) are respectively given by

\[D_L(m,n) = D_{max}(m,n) + D_e(m,n) + D_g(m,n) \]  \hspace{1cm} (6)

\[D_H(m,n) = D_{min}(m,n) + bD_g(m,n) \]  \hspace{1cm} (7)

where \(0 < b < 1\) serves as a lower bound of spatial constraint. By combining the low and high complexity distance and complexity level factor \(\rho_{m,n}\), we define the complexity-adaptive distance as

\[D_{total} = \rho_{m,n}D_L + (1 - \rho_{m,n})D_H + \eta D_S \]  \hspace{1cm} (8)

**Table 1:** Comparison results of MABO and AUC using 500, 1000, and 2000 candidates; CA1 and CA2 respectively are the two settings used in our method.

<table>
<thead>
<tr>
<th>Methods</th>
<th>500 Candidates</th>
<th>1000 Candidates</th>
<th>2000 Candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MABO</td>
<td>AUC</td>
<td>MABO</td>
</tr>
<tr>
<td>Selective Search [4]</td>
<td>0.771</td>
<td>0.517</td>
<td>0.799</td>
</tr>
<tr>
<td>MCG [2]</td>
<td>0.757</td>
<td>0.510</td>
<td>0.782</td>
</tr>
<tr>
<td>EdgeBox [5]</td>
<td>0.755</td>
<td>0.520</td>
<td>0.782</td>
</tr>
<tr>
<td>SPA [3]</td>
<td>0.736</td>
<td>0.487</td>
<td>0.776</td>
</tr>
<tr>
<td>CA1</td>
<td>0.768</td>
<td>0.517</td>
<td>0.809</td>
</tr>
<tr>
<td>CA2</td>
<td>0.775</td>
<td>0.536</td>
<td>0.812</td>
</tr>
</tbody>
</table>

Figure 1: Color distance metrics. (a) The mean color is well behaved to delineate low-complexity superpixels, but (b) such aggregate measure fails to reflect the distance between two high-complexity superpixel sets.