Using Multiple Segmentations to Discover Objects and their Extent in Image Collections

Bryan C. Russell, Alexei A. Efros, Josef Sivic, William T. Freeman and Andrew Zisserman

Presented by Priyanka Godbole
4/10/2007

Goal

- Given a collection of unlabelled images, discover visual object categories and their segmentation.

Approach

The multiple segmentation approach to automatically discover objects and thus the visual category uses a 4 step algorithm:

- Normalized Cuts algorithm - compute multiple segmentations for each image.
- For each segment compute a histogram of visual words.
- Perform topic discovery on all segments, treating each segment as a document.
- Sort the segments based on similarity between visual words of segment and topic.

Why multiple Segmentation?

- Choice of a seg. algo. NOT CRITICAL
- We do not rely on FULL segmentation to be correct
- NOT STABLE as the output changes when the parameters are changed.
- So we use a multiple seg. approach

Normalized Cuts Algorithm

- It produces a global segmentation such that large segments could be objects.
- To produce multiple segmentations, vary 2 parameters –
  - # of segments K (K = 3, 5, 7, 9)
  - Size of input image (2 image scales – 50 and 100 pixels across)
  - For LabelMe dataset, K= 11, 13 also used
  - For MSRC dataset, image scale = 150 pixel across also used.

Multiple segmentations

We use Normalized Cuts, varying parameter settings: # segments and image scale
Intuitions

- Intuition #1: All segmentations are wrong, but some segments are good.
- Intuition #2: All good segments are alike, each bad segment is bad in its own way.

Obtaining visual words

- Due to imperfection in segmentation → representation is tolerant to partial occlusion and clutter.

Bag-of-words Approaches

- Represent image as a histogram of visual words.
- Detect affine covariant regions.
- Represent each region by a SIFT descriptor.
- Build visual vocabulary by k-means clustering (K~1,000).
- Assign each region to the nearest cluster centre.

Topic discovery

- It partitions the segmented objects into visual object classes.

It uses:
- probabilistic Latent Semantic Analysis (pLSA).
- Latent Dirichlet Allocation (LDA).
- It uses unordered "bag of words" representation of documents.

Topic Discovery

- Representing Segments
- Finding coherent segment clusters (topics):
  \( w \) ... visual words
  \( d \) ... documents (images)
  \( z \) ... topics (‘objects’)

Use statistical text analysis techniques such as Latent Semantic Analysis (LSA), Probabilistic LSA (Hofmann ’99), or Latent Dirichlet Allocation (LDA) [Blei et al. ’03]. Here we chose LDA.
Visual word shortcomings - 1

- **Visual Synonyms**: Two different visual words representing a similar part of an object (wheel of a motorbike).

Visual word shortcomings - 2

- **Visual Polysemy**: Single visual word occurring on different (but locally similar) parts on different object categories.

Visual word shortcomings - 3

- **Lack of hard segmentation**

Segment Scoring

Compare segment distributions against learned topic distribution over visual words using KL divergence.

Results

**Retrieval Accuracy**: Average precision for MSRC

- For bicycles and windows the method performs on par or better than the other methods.
- Precision recall curves are evaluated and average precision is reported.

<table>
<thead>
<tr>
<th>Method</th>
<th>bicycles</th>
<th>cars</th>
<th>signs</th>
<th>windows</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Mult. seg. LDA</td>
<td>0.69</td>
<td>0.77</td>
<td>0.43</td>
<td>0.74</td>
</tr>
<tr>
<td>(b) Mult. seg. j,s,SA</td>
<td>0.67</td>
<td>0.28</td>
<td>0.34</td>
<td>0.57</td>
</tr>
<tr>
<td>(c) Sing. seg. LDA</td>
<td>0.67</td>
<td>0.73</td>
<td>0.46</td>
<td>0.72</td>
</tr>
<tr>
<td>(d) No seg. LDA</td>
<td>0.64</td>
<td>0.85</td>
<td>0.40</td>
<td>0.74</td>
</tr>
<tr>
<td>(e) Chance</td>
<td>0.06</td>
<td>0.12</td>
<td>0.04</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Results

Segmentation Accuracy: Average overlap area score for Label Me

The LabelMe dataset is more difficult as the images are taken in the natural habitat.

<table>
<thead>
<tr>
<th>Method</th>
<th>buildings</th>
<th>cars</th>
<th>roads</th>
<th>sky</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Mult. seg. LDA</td>
<td>0.53</td>
<td>0.21</td>
<td>0.41</td>
<td>0.77</td>
</tr>
<tr>
<td>(b) Mult. seg. pLSA</td>
<td>0.59</td>
<td>0.09</td>
<td>0.16</td>
<td>0.77</td>
</tr>
<tr>
<td>(c) Sing. seg. LDA</td>
<td>0.55</td>
<td>0.29</td>
<td>0.32</td>
<td>0.65</td>
</tr>
<tr>
<td>(d) No. seg. LDA</td>
<td>0.47</td>
<td>0.16</td>
<td>0.14</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Results: Top segments Montages

Caltech 5
10 topics, 4090 images

Results: Top segments Montages

MSRC Set
25 topics, 4325 images

Results: Top segments Montages

Label Me
20 topics, 1554 images

Thank You