Shape Discovery from Unlabeled Image Collections

Yong Jae Lee and Kristen Grauman
University of Texas at Austin

Problem
Can we discover common object shapes within unlabeled multi-category collections of images?

Main idea
We propose a method that analyzes a collection of unlabeled images and returns:
1) a set of proposed prototypical shape models
2) per-image estimates of foreground contours

Algorithm Overview
Anchoring Edge Fragments to Local Patches
Since edge features often lack distinctiveness, we use patch appearance matches (e.g., SIFT) to initialize regions for shape matching.

Computing Feature Matches with Anchored Edge Features
Compare each feature in X for all features in Y
Align X’s edgenet onto Y’s edgenet based on the best matching feature
Each fragment in X is fitted to the nearest best matching fragment in Y

d_{\text{match}}(f_i, f_j, f'_k)
d_{\text{shape}}(f_i, f_j, f'_k)

* Dotted circles represent the Gaussian weighting on the edge fragments with the highest weight at feature center.

Grouping Cluttered Images with Similar Shapes
The overall directed patch and shape distance from image X to image Y is the average over the component feature distances between each f_i, j in X and its best matching feature f_{i,j} in Y:

\[ d_{\text{match}}(X, Y) = \frac{1}{|X|} \sum_{i,j} d_{\text{match}}(f_i, j, f_{i,j}) \quad d_{\text{shape}}(X, Y) = \frac{1}{|X|} \sum_{i,j} d_{\text{shape}}(f_i, j, f_{i,j}). \]

We obtain a symmetric cost via the sum:

\[ D(X, Y) = D(X, Y) + D(Y, X). \]

Inferring Foreground Contours

Prototypical Shape Formation
Not all images in a cluster will contain the same object or agree in terms of shape.

→ Create a simple vote space with the computed edge weights such that the common shape can be reinforced in the output while parts that agree less can be discarded.

Results
We present results to analyze our method’s unsupervised category and shape discovery using the Caltech, ETHZ shape, and LabelMe datasets.

Unsupervised Category Discovery
<table>
<thead>
<tr>
<th>Category</th>
<th>Caltech</th>
<th>ETHZ</th>
<th>LabelMe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals</td>
<td>95.5%</td>
<td>77.6%</td>
<td>65.6%</td>
</tr>
<tr>
<td>Art</td>
<td>98.4%</td>
<td>77.6%</td>
<td>65.6%</td>
</tr>
<tr>
<td>Objects</td>
<td>94.5%</td>
<td>77.6%</td>
<td>65.6%</td>
</tr>
<tr>
<td>Vehicles</td>
<td>90.3%</td>
<td>77.6%</td>
<td>65.6%</td>
</tr>
</tbody>
</table>

* Lower curves are better

 Foreground Shape Discovery: Localization

- Bounding Box Hit Rate (BBHR) measures the % of images in the dataset that have at least h = 5 foreground features selected, as a function of the selection threshold applied to the feature weights.
- False Positive Rate (FPR) measures the average % of selected features falling outside of the bounding box.

Our method outperforms a state-of-the-art method and achieves very good localization rates.

Foreground Shape Discovery: Prototypical Shape

Algorithm Flow Chart: Recap

* Dotted circles represent the Gaussian weighting on the edge fragments with the highest weight at feature center.
**Foreground Shape Discovery: Prototypical Shape**

- ETHZ Images (bounding box)
- ETHZ Images (expanded)

**Generalization to Detection in Novel Images**

- Detection task on LabelMe dataset
- Match prototypical shapes to the test images using modified chamfer matching
  - $\alpha_2 = (I(H_{g1}\cap H_{g2})|I(H_{g1}\cup H_{g2}))$
  - Average $\alpha_2$: [F: 0.47, A: 0.43, M: 0.38, C: 0.31]
  - Chance Detection: [F: 0.03, A: 0.02, M: 0.03, C: 0.02]

**Conclusions**

- We have proposed the first method to discover common object shapes within unlabeled multi-category collection of images.
- We have shown the strength of our patch-anchored shape matching by comparing against baseline methods that use each feature in isolation, as well as against previous appearance-only unsupervised learners.