Learning a Classification Model for Segmentation

Segmentation as Classification

What is a Good Segmentation?

- Elements inside one region are similar:
  - Similar brightness
  - Similar texture
  - Weak contours in interior
- Elements in different regions are dissimilar:
  - Dissimilar brightness
  - Dissimilar texture
  - Strong contours along region boundaries
- Curvilinear continuity:
  - Smooth boundaries

Features for Classification

- Intra-region similarity
  - Brightness similarity
  - Texture similarity
- Inter-region similarity
  - Brightness similarity
  - Texture similarity
  - Intra-region contour energy
  - Inter-region contour energy
  - Curvilinear continuity

Procedures

- Preprocessing – Partition the pixels to the superpixels
- Features - define the features
- Classifier – how to combine them using a simple linear classifier
- Search – MCMC based search algorithm

Superpixels

- Pixels are not natural entities.
- The number of pixels is high.
- Superpixels are local, coherent and which preserves most of the structure necessary for segmentation.

![Figure 1](image1.png) Figure 1. We formulate segmentation as classification between good segmentations (b) and bad segmentations (c). We use Gestalt grouping cues as features and train a classifier. Human segmented images are used as examples of good segmentations. Bad segmentations are constructed by randomly matching a human segmentation to a different image.

![Figure 2](image2.png) Figure 2. An example of superpixel maps. (a) is the original image. (b) is a human marked segmentation; (c) is a superpixel map with \( r = 0.3 \). (d) shows a reconstruction of the human segmentation from the superpixels; we assign each superpixel to a segment as in (b) with the regions overlapping areas and extract the superpixel boundaries.
Preprocessing: Pixels to Superpixels

- Use normalized cut algorithm to make superpixels.
- The criterion for partitioning the graph
  - minimize the sum of weights of connections across the groups.
  - maximize the sum of weights of connections within the groups.

Texton

- The representation of textures using filter responses is redundant.
- Textures with some repeating properties.
- Clustering the filter responses into a small set of prototype response vectors (textons) is needed.
  - The image is convolved with a bank of filters of multiple orientations.
  - Based on the filter output, the pixels are clustered into a number of texton channels.
  - The resulting distribution of textons for each region makes histograms.

Texton

- The representation of textures using filter responses is redundant.
- Textures with some repeating properties.
- Clustering the filter responses into a small set of prototype response vectors (textons) is needed.
  - The image is convolved with a bank of filters of multiple orientations.
  - Based on the filter output, the pixels are clustered into a number of texton channels.
  - The resulting distribution of textons for each regions makes histograms.

Texton

- The representation of textures using filter responses is redundant.
- Textures with some repeating properties.
- Clustering the filter responses into a small set of prototype response vectors (textons) is needed.
  - The image is convolved with a bank of filters of multiple orientations.
  - Based on the filter output, the pixels are clustered into a number of texton channels.
  - The resulting distribution of textons for each regions makes histograms.

Texture Similarity

- The texture difference of two regions is measured as the $X^2$ distance between two histograms.

Texture Similarity

- The intra-region similarity compares the descriptor of a superpixel $q$ to the segment $S$ containing it.
- The inter-region similarity compares the descriptor of a superpixel $q$ on the boundary of $S'$ to the adjacent segment.

Contour Energy

- The oriented energy at angle $0$ is defined as
  $$E_0 = ||1 + y_1||^2 + ||1 + y_2||^2$$
Contour Energy
- Intra-region contour energy is the average orientation energy on the superpixel boundaries on the interior of S.
- Inter-region contour energy is the average orientation energy on the boundary of S.

Good Continuation
- Curvilinear continuity of S is the average of tangent changes for all pairs of superpixels on the boundary of S.

Power of the Gestalt Cues

<table>
<thead>
<tr>
<th>Feature</th>
<th>Information</th>
<th>Residual Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contour</td>
<td>0.387</td>
<td>0.010</td>
</tr>
<tr>
<td>Texture</td>
<td>0.123</td>
<td>0.016</td>
</tr>
<tr>
<td>Texture</td>
<td>0.137</td>
<td>0.005</td>
</tr>
<tr>
<td>Texture</td>
<td>0.016</td>
<td>0.006</td>
</tr>
<tr>
<td>Brightness</td>
<td>0.117</td>
<td>0.005</td>
</tr>
<tr>
<td>Brightness</td>
<td>0.016</td>
<td>0.007</td>
</tr>
<tr>
<td>Continuity</td>
<td>0.191</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Training the classifier
- Use a simple logistic regression classifier

\[ G(S) = \sum_j c_j F_j(S) - \theta \]

The higher the value of G is, the more likely S is a good segment.

Finding good segmentations
- It becomes the optimization of f in the space of all segmentations.

\[ f(S) = \sum_c \left( \sum_j c_j F_j(S) - \theta \right) \]
- The search space is large, so do the random search.
Search for Good Segmentation

- Linear objective function
- At each step, randomly construct a new segmentation, based on simulated annealing.
- Local search dynamics involves three basic moves.
  - Shift
  - Merge
  - Split

Conclusion

- It treats the segmentation as the classification of good and bad segmentations.
- The Gestalt grouping cues are combined in a principled way.
- A linear classifier and a simple random search algorithm.
- Still difficult optimization problem.