Learning a Classification Model for Segmentation

Segmentation as Classification

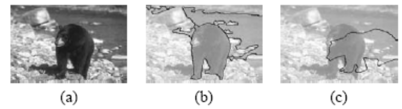


Figure 1. We formulate segmentation as classification between good segmentations (b) and bad segmentations (c). We use Gestalt grouping cues as features and train a classifier. Human segmented images are used as examples of good segmentations. Bad segmentations are constructed by randomly matching a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

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
- Cuvilinear 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.

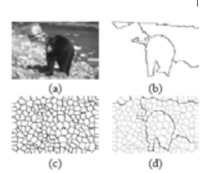


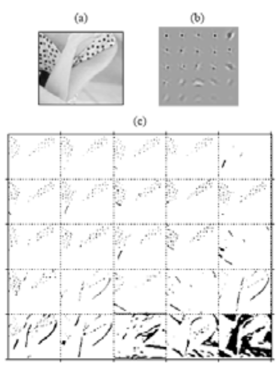
Figure 2. An example of superpixel maps. (a) is the original image; (b) is a human marked segmentation; (c) is a superpixel map with $k = 200$; (d) shows a reconstruction of the human segmentation from the superpixels: we assign each superpixel to a segment in (b) with the maximum overlapping area and 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

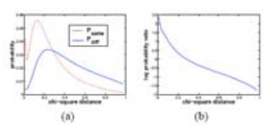
- The representation of textures using filter responses is redundant.
- Textures with some repeating properties.
- Clustering the filter responses into a small set of prototype response vectors (textons) is needed.
  - The image is convolved with a bank of filters of multiple orientations.
  - Based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



Texture Similarity

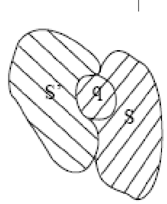
- The texture difference of two regions is measured as the X^2 distance between two histograms.

$$T(q,S) = \log \frac{P_{same}(d_T(q,S))}{P_{diff}(d_T(q,S))}$$



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 angel 0 is defined as

$$OE_{0^\circ} = (I * f_1)^2 + (I * f_2)^2$$

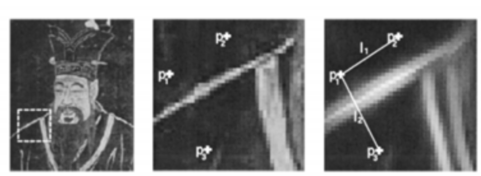


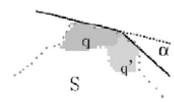
Figure 8. Left: the original image. Middle: part of the image marked by the box. The intensity values at pixels p_1, p_2 and p_3 are similar. However, there is a contour in the middle, which suggests that p_1 and p_2 belong to one group while p_3 belongs to another. Just comparing intensity values at these three locations will mistakenly suggest that they belong to the same group. Right: orientation energy. Somewhere along l_2, the orientation energy is strong which correctly proposes that p_1 and p_3 belong to two different partitions, while orientation energy along l_1 is weak throughout, which will support the hypothesis that p_1 and p_2 belong to the same group.

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.



(b)

Power of the Gestalt Cues

Feature		Information	Residual Info.
Contour:	inter-	0.387	0.010
	intra-	0.012	0.010
Texture:	inter-	0.137	0.005
	intra-	0.030	0.008
Brightness:	inter-	0.112	0.005
	intra-	0.049	0.007
Continuity:		0.198	0.002

(a)

Combined Feature	Information	Residual Info.
Contour	0.510	0.024
Texture	0.220	0.026
Brightness	0.232	0.025

(b)

Training the classifier

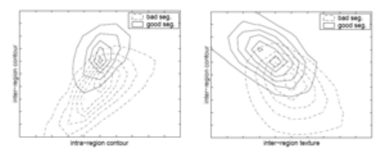


Figure 6. Iso-probability contour plots of empirical distributions for a pair of features. These plots suggest that: (1) the normalized features are well-behaved; for both classes a Gaussian model would be a reasonable approximation. And (2) a linear classifier would perform well.

Training the classifier

- Use a simple logistic regression classifier

$$\mathbf{G}(S) = \sum_j c_j \mathbf{F}_j(S) - \theta$$

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

Finding good segmentations

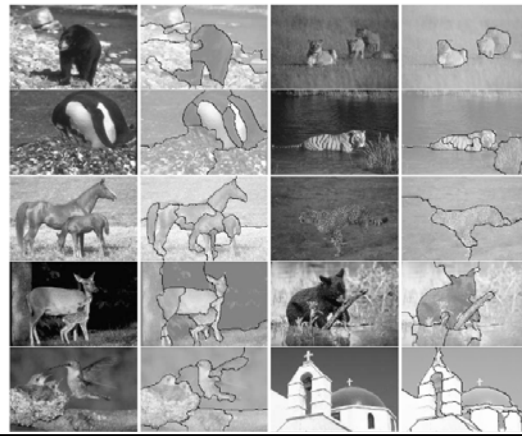
- It becomes the optimization of f in the space of all segmentations.

$$f(\mathcal{S}) = \sum_{S \in \mathcal{S}} \left(\sum_j c_j \mathbf{F}_j(S) - \theta \right)$$

- 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.