Superpixel: Empirical Studies and Applications

Introduction

Many existing algorithms in computer vision use the pixel-grid as the underlying representation. For example, stochastic models of images, such as Markov random fields, are often defined on this regular grid. Or, face detection is typically done by matching stored templates to every fixed-size (say, 50x50) window in the image.

The pixel-grid, however, is **not** a natural representation of visual scenes. It is rather an "artifact" of a digital imaging process. It would be more natural, and presumably more efficient, to work with perceptually meaningful entities obtained from a low-level grouping process. For example, we can apply the Normalized Cuts algorithm to partition an image into, say, 500 segments (what we call **superpixels**).

Such a superpixel map has many desired properties:

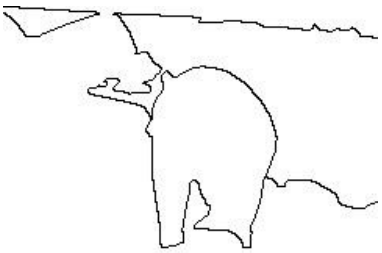
- It is **computationally efficient**: it reduces the complexity of images from hundreds of thousands of pixels to only a few hundred superpixels.
- It is also **representationally efficient**: pairwise constraints between units, while only for adjacent pixels on the pixel-grid, can now model much longer-range interactions between superpixels.
- The superpixels are **percetually meaningful**: each superpixel is a perceptually consistent unit, i.e. all pixels in a superpixel are most likely uniform in, say, color and texture.
- It is **near-complete**: because superpixels are results of an oversegmentation, most structures in the image are conserved. There is very little loss in moving from the pixel-grid to the superpixel map.

It is actually not novel to use superpixels or atomic regions to speed up later-stage visual processing; the idea has been around the community for a while. What we have done is: (1) to empirically validate the completeness of superpixel maps; and (2) to apply it to solve challenging vision problems such as finding people in static images.

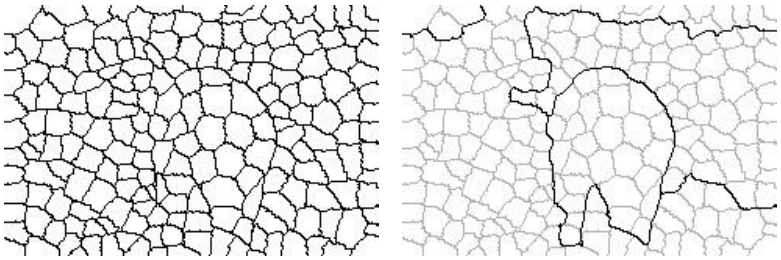
Superpixels from the Normalized Cuts

The Normalized Cuts is a classical region segmentation algrithm developed at Berkeley, which uses spectral clustering to exploit pairwise brightness, color and texture affinities between pixels. We apply the Normalized Cuts to oversegment images to obtain superpixels. In our experiments, to enforce locality we use only local connections in the pairwise affinity matrix.

(a)



(b)



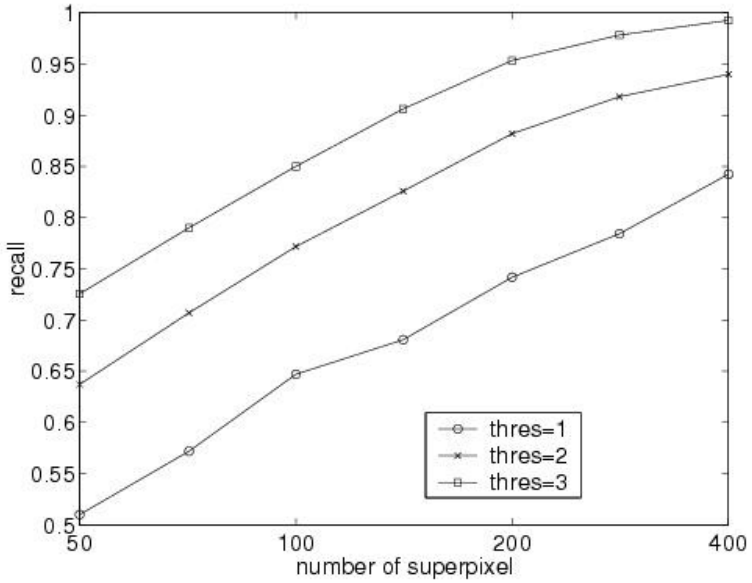
(c) (d)

Figure 1: 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.

Figure **1** shows an example of superpixels. If we compare the human-marked segmentation (the groundtruth) to the one reconstructed from the superpixel map, we may find that some contour details are lost in the process of oversegmentation (such as at the upper-left corner). However most structures are conserved; and the reconstructed segmentation is qualitatively very similar to the groundtruth.

Empirical Validation

To empirically validate the completeness of superpixels maps, we use a boundary-based strategy: each human-marked segmentation defines a boundary map (Figure 1(b)); so does each superpixel map (Figure 1(c)). We can try to match the groundtruth boundaries to the superpixel boundaries: if a groundtruth boundary pixel is within a distance threshold (say, 2 pixels) of a superpixel boundary pixel, we count it as a hit. The overall **recall rate**, percentage of groundtruth boundary pixels being matched to superpixel boundaries, is a measure of how much structure is being conserved.



For images of size (240x160) in the BSDS, the figure above shows the recall rates of superpixel maps from k=50 to 400. Overall the recall rates are quite good (especially if taken into consideration the complexities of the BSDS images), indicating that a state-of-

the-art segmentation algorithm is able to produce superpixel maps with little loss of perceptually important structure. In particular, at the threshold of 2 pixels, the recall rate is about 90% for 200 superpixels. This is sufficient for our work in learning discriminative models of segmentation and finding people in static images.

Greg Mori has released a version of our superpixel code in matlab. Alyosha Efros has used other region segmentation algorithms in his recent work using superpixels. Alternatively, we have developed a boundary-oriented superpixel algorithm, the CDT graphs, which is scale-invariant (and very fast).

References

1. **Learning a Classification Model for Segmentation**. [abstract] [pdf] [ps] [bibtex]
   Xiaofeng Ren and Jitendra Malik, in *ICCV '03*, volume 1, pages 10-17, Nice 2003.
2. **Recovering Human Body Configurations: Combining Segmentation and Recognition**. [abstract] [pdf] [ps] [bibtex]
   Greg Mori, Xiaofeng Ren, Alyosha Efros and Jitendra Malik, in *CVPR '04*, volume 2, pages 326-333, Washington, DC 2004.