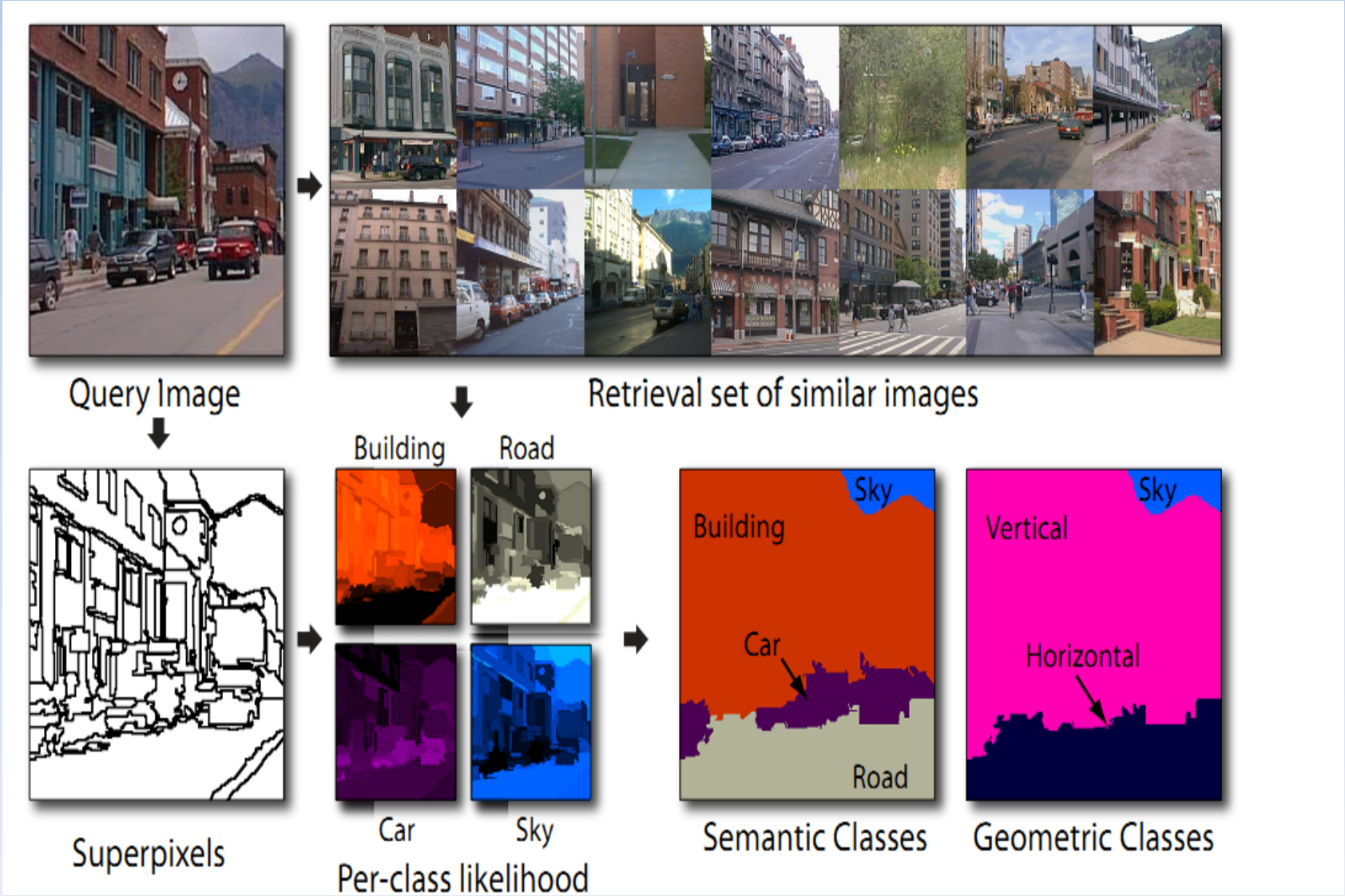
Image Parsing

- Joseph Tighe & Svetlana Lazebnik
- ECCV ‘10


Query Image
Retrieval set of similar images
Superpixels
Building
Road
Car
Sky
Per-class likelihood
Sky
Building
Car
Road
Semantic Classes
Sky
Vertical
Horizontal
Geometric Classes

What is Image Parsing

- simple and efective nonparametric approach to the problem of image parsing, namely labeling image regions
- Make use of global scene-level matching, superpixel-level matching and Markov random eld (MRF) optimization.

Challenges

- the dataset is constantly expanding as people upload new photos
- Training can take days
- We want to work with large datasets
- Also , SIFT ow is fairly complex and expensive to solve.

Retrieval Set

- retrieval set of scenes whose content is used to interpret the test image
- retrieval set images are very similar to the test image in terms of spatial layout of the classes, so that we transfer labels at the level of superpixels

Relationship with Context

- simultaneously estimate a semantic label (e.g., building,car, person, etc.) and a geometric label (sky, ground,or vertical surface) while making sure the two types of
- labels assigned to the same region are consistent

Method Summary

- 1. Find a retrieval set of images similar to the query image
- 2. Segment the query image into superpixels and compute feature vectors for each superpixel
- 3. For each superpixel and each feature type, nd the nearest-neighbor superpixels in the retrieval set according to that feature. Compute a likelihood score for each class based on the superpixel matches .

Method Summary

- 4. Use the computed likelihoods together with pairwise co-occurrence energies in an Markov Random Field (MRF) framework to compute a global labeling of the image . Alternatively, with modifications , the MRF framework can simultaneously solve for both semantic and geometric class labels

Global Features vs. Superpixel Features

- Global : Spatial pyramid,Gist, Color histogram
- (To decide the retreival set)
- Superpixel : Shape,Location,Texture,Color,Appearance
- (To compute the closeness)

Math

$$L(s_i, c) = \log \frac{P(s_i|c)}{P(s_i|\bar{c})} = \log \prod_k \frac{P(f_i^k|c)}{P(f_i^k|\bar{c})}$$
$$= \sum_k \log \frac{P(f_i^k|c)}{P(f_i^k|\bar{c})}, \quad (1)$$

$$\frac{P(f_i^k \mid c)}{P(f_i^k \mid \bar{c})} = \frac{(n(c, \mathcal{N}_i^k) + \epsilon)/n(c, \mathcal{D})}{(n(\bar{c}, \mathcal{N}_i^k) + \epsilon)/n(\bar{c}, \mathcal{D})}$$
$$= \frac{n(c, \mathcal{N}_i^k) + \epsilon}{n(\bar{c}, \mathcal{N}_i^k) + \epsilon} \times \frac{n(\bar{c}, \mathcal{D})}{n(c, \mathcal{D})}, \quad (2)$$

Interpretations

- Likelihood of superpixel i to the class c : independent multiplication in the neigborhood of f.

Math(cont.)

$$J(\mathbf{c}) = \sum_{s_i \in SP} E_{\text{data}}(s_i, c_i) + \lambda \sum_{(s_i, s_j) \in A} E_{\text{smooth}}(c_i, c_j)\,, \tag{3}$$

$$E_{\text{data}}(s_i, c_i) = -w_i \sigma(L(s_i, c_i)), \tag{4}$$

$$E_{\text{smooth}}(c_i, c_j) = -\log[(P(c_i|c_j) + P(c_j|c_i))/2] \times \delta[c_i \neq c_j]\,, \tag{5}$$

$$H(\mathbf{c}, \mathbf{g}) = J(\mathbf{c}) + J(\mathbf{g}) + \mu \sum_{s_i \in SP} \varphi(c_i, g_i), \tag{6}$$

Interpretations

- Data and smoothness parameters to minimize,
- given a weight to superpixels

- Smoothness parameter is defined based on probabilities of label co-occurence

Results

- boosted decision tree outputs a likelihood ratio score that is comparable to the one produced by nonparametric scheme but that gets about 2% higher accuracy for geometric classification
- MRF improves the results for both types of classes
- Stuff (people,car) results are better than Things (sky,road)

Results

Final system achieves a classification rate of 54.9% across all scene types

As for joint semantic/geometric inference, it not only gives the highest overall accuracy in all cases, but is also much less prone to over-smoothing.