SuperParsing: Scalable Nonparametric Image Parsing with Superpixels

Joseph Tighe and Svetlana Lazebnik
Dept. of Computer Science, University of North Carolina at Chapel Hill
http://www.cs.unc.edu/SuperParsing

Overview

- **"Open universe" system:** no training required, easy to accommodate an evolving dataset as new classes or new training exemplars are added
- **State of the art performance** on the SIFT Flow dataset (Liu et al., CVPR 2009)
- **New large-scale baseline for image parsing:** per-pixel recognition results on a subset of LabelMe consisting of 15k images, 170 labels

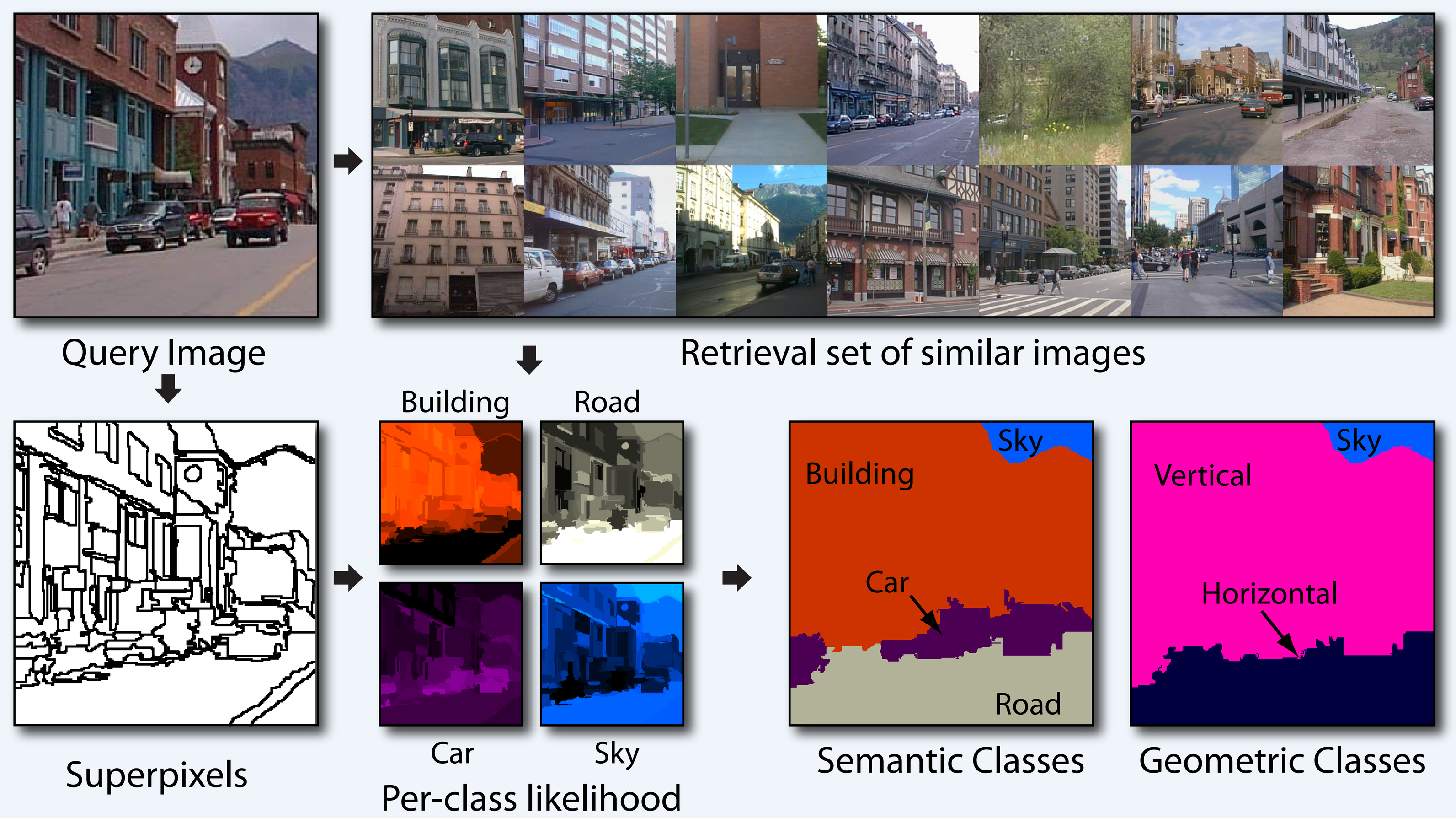


Image Parsing Method

Given a query (test) image:

- Find a retrieval set of 200 similar images by taking the minimum per-feature rank of four global image features.
- For each test superpixel s_i described by multiple local features $\{f_i^k\}$, compute a likelihood ratio score for each class c found in the retrieval set:

$$L(s_i, c) = \frac{P(s_i|c)}{P(s_i|\neg c)} = \prod_k \frac{P(f_i^k|c)}{P(f_i^k|\neg c)} .$$

The feature likelihood $P(f_i^k|c)$ is given by a nonparametric density estimate:

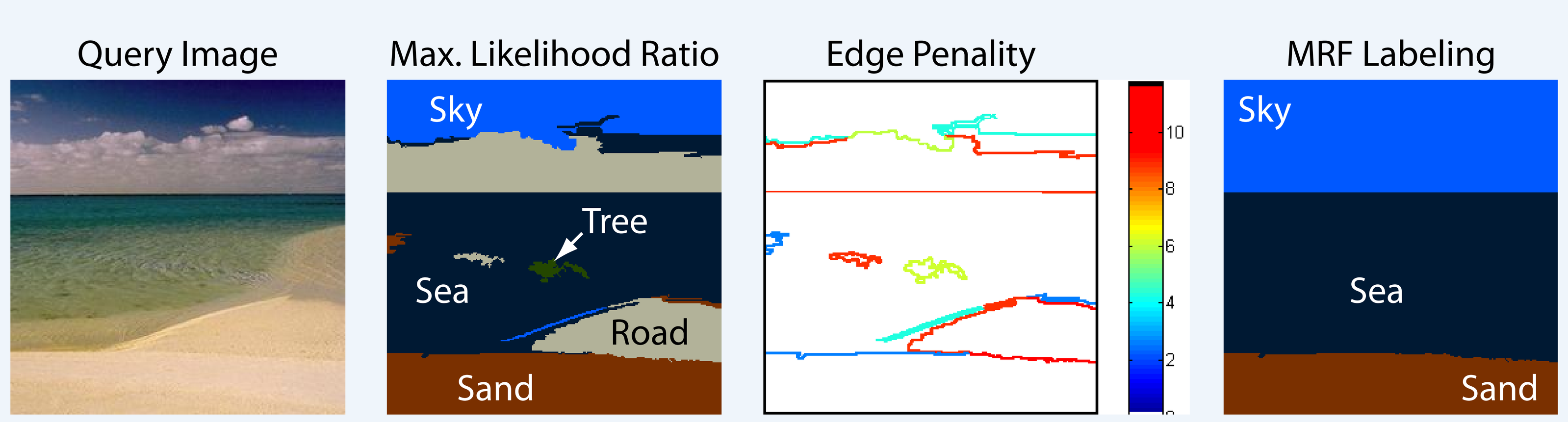
$$P(f_i^k \mid c) = \frac{\#(\text{retrieval set features of class } c \text{ within a fixed radius of } f_i^k)}{\#(\text{total features of class } c \text{ in the training set})} .$$

- Use MRF inference to solve for the label field $\mathbf{c} = \{c_i\}$ over the entire test image:

$$J(\mathbf{c}) = \sum_{s_i \in SP} -w_i \log L(s_i, c_i) + \lambda \sum_{(s_i,s_j)\in A} E_{\text{edge}}(c_i, c_j) ,$$

where w_i is a weight based on the superpixel size, and edge penalty E_{edge} is based on the co-occurrence of adjacent labels in the training set:

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

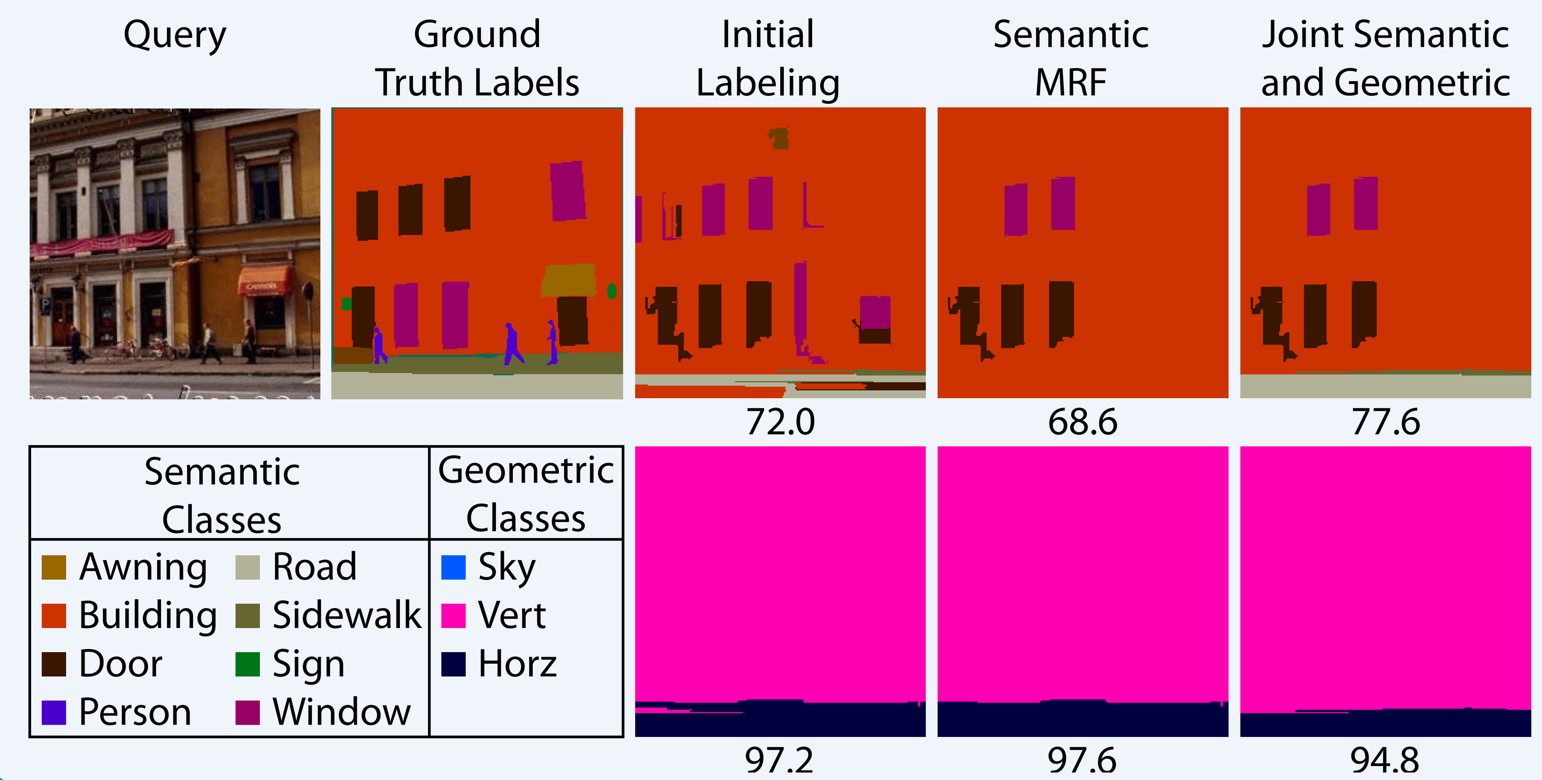


Joint Semantic and Geometric Labeling

Simultaneously solve for a field of *semantic* labels (c) and *geometric* labels (g) over the image by optimizing

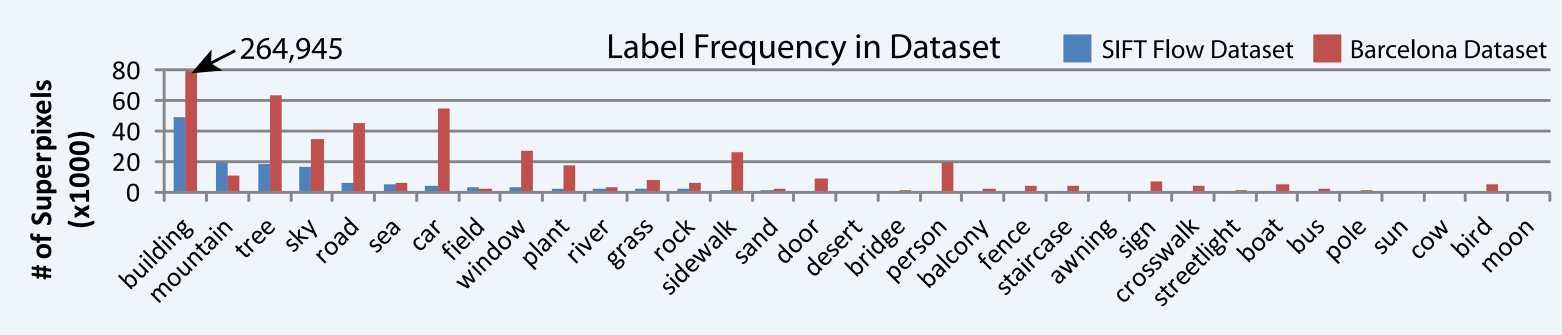
$$H(\mathbf{c}, \mathbf{g}) = J(\mathbf{c}) + J(\mathbf{g}) + \mu \sum_{s_i \in SP} \varphi(c_i, g_i) ,$$

where $\varphi(c_i, g_i)$ is a coherence term between the semantic and geometric label of the same superpixel.



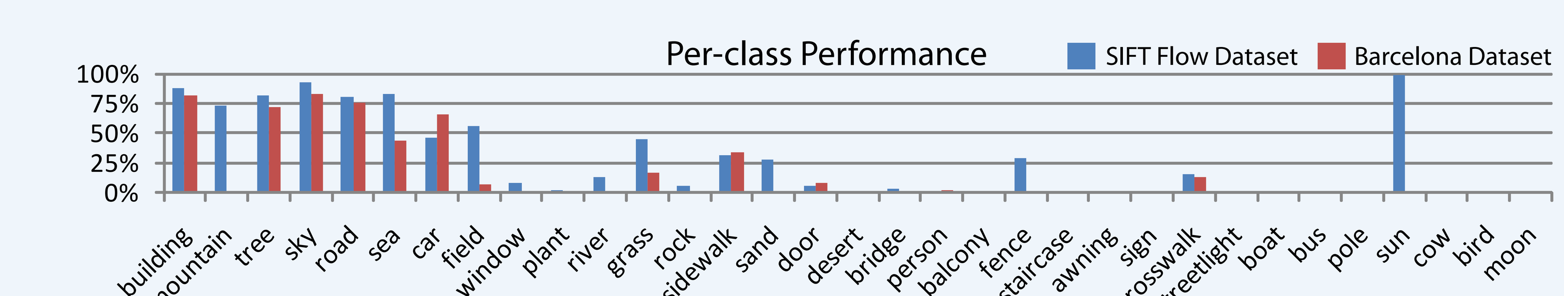
Results on Large-Scale Datasets

- **SIFT Flow dataset** (Liu et al., CVPR 2009): 2,488 training images, 200 test images, 33 labels
- **Barcelona dataset** (a new large-scale benchmark): 14,871 training images, 279 test images, 170 labels



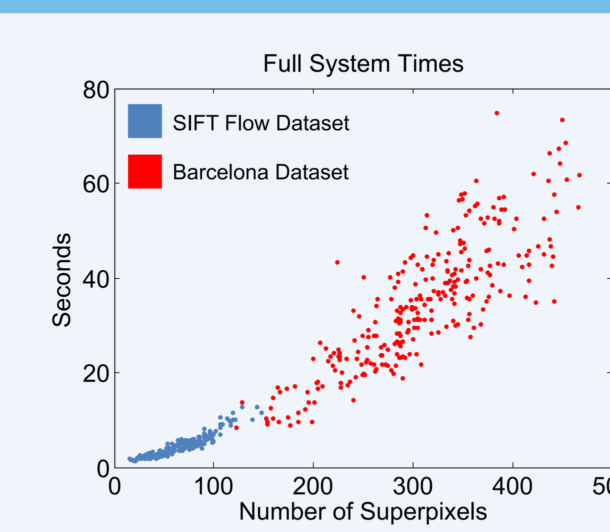
Per-pixel classification rates (with average per-class rates in parentheses):

	SIFT Flow		Barcelona	
	Semantic	Geometric	Semantic	Geometric
Liu et al. (CVPR 2009)	74.75	N/A	N/A	N/A
Local labeling	73.2 (29.1)	89.8	62.5 (8.0)	89.9
MRF	76.3 (28.8)	89.9	66.6 (7.6)	90.2
Joint semantic/geometric	76.9 (29.4)	90.8	66.9 (7.6)	90.7

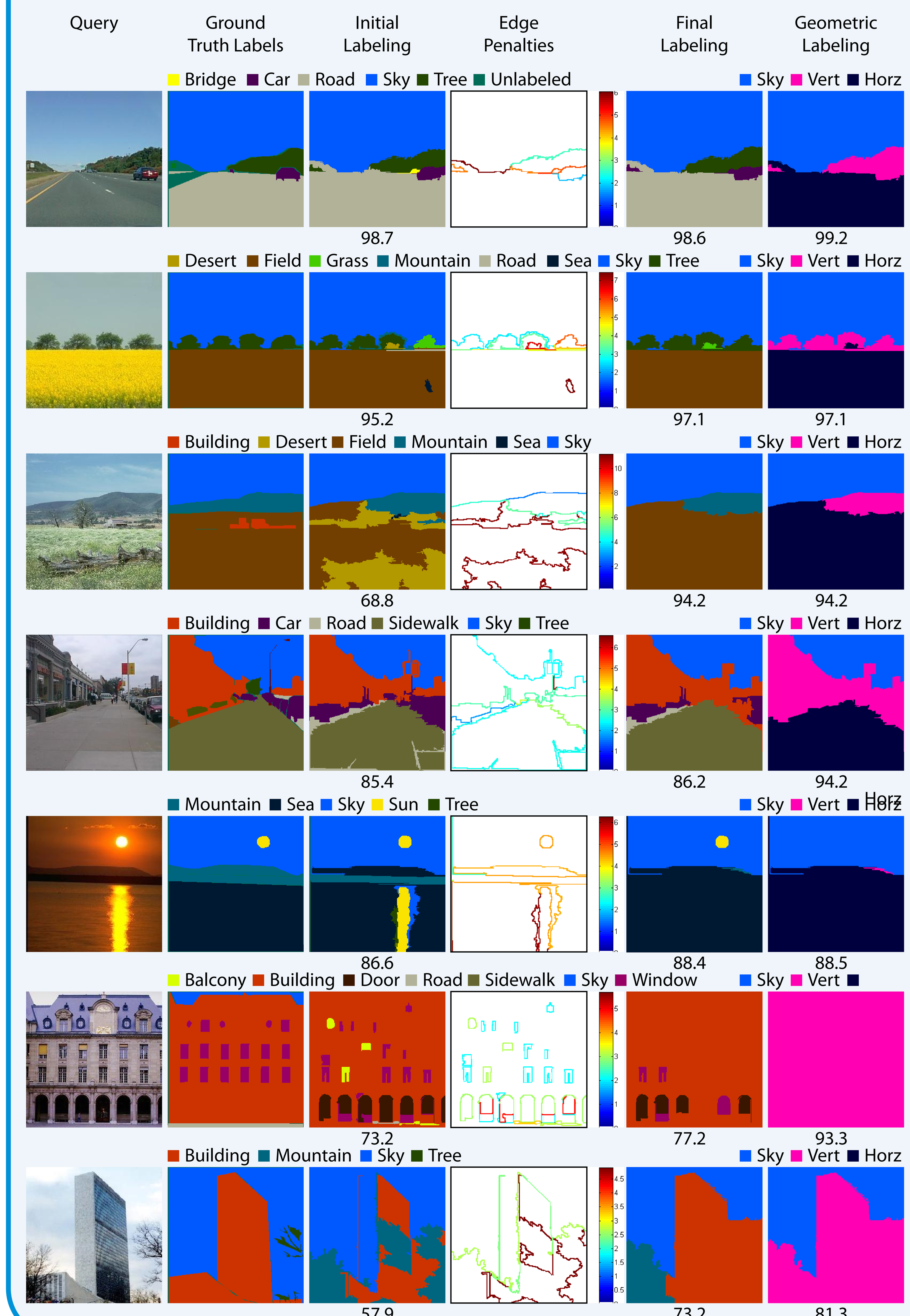


Timing

	SIFT Flow	Barcelona
Training set size	2,488	14,871
Image size	256 × 256	640 × 480
Ave. # superpixels	63.9	307.9
Feature extraction	∼ 4 sec	∼ 5 min
Retrieval set search	0.04 ± 0.0	0.21 ± 0.0
Superpixel search	4.4 ± 2.3	34.2 ± 13.4
MRF solver	0.005 ± 0.003	0.03 ± 0.02
Total (excluding features)	4.4 ± 2.3	34.4 ± 13.4



Sample Output on SIFT Flow Dataset



Small Datasets

Results on two small-scale datasets using trained boosted decision tree classifiers (instead of retrieval set and superpixel search):

	Stanford Dataset 715 images, 8 classes		Geometric Context Dataset 300 images, 7 classes	
	Semantic	Geometric	Sub-classes	Main classes
Gould et al. (ICCV 2009)	76.4	91.0	N/A	86.9
Hoiem et al. (IJCV 2007)	N/A	N/A	61.5	88.1
Local labeling	76.9	90.5	57.6	87.8
MRF	77.5	90.6	61.0	88.2
Joint semantic/geometric	77.5	90.6	61.0	88.1

Funding

This research was supported in part by NSF CAREER award IIS-0845629, Microsoft Research Faculty Fellowship, and Xerox.